

Rethink anti-bioterrorism plans

With the US National Academies about to pronounce on science's role in counteracting bioterrorists, it falls on the Congress to oppose and restructure the domestic security framework proposed by the Bush administration.

In the immediate aftermath of the events of 11 September 2001, the US public's reaction to the calamity that had befallen their country was characterized by remarkable calm and determination.

Despite the grief engendered by the terrorist attacks and the great difficulties that appeared before the nation last autumn, most people expressed strong faith that the country and its institutions could steer through the challenges ahead. This belief carried with it considerable, perhaps unreasonable, expectations. One of these — implicit in the first images of precision bombing of the bedraggled foe in Afghanistan — was that the United States' formidable expertise in science and technology would help it to navigate these challenges.

Leading scientists, mindful of their prominent role in the Second World War and the Cold War, were quick to volunteer their services in the new 'war against terrorism'. The National Academies of science and engineering and the Institute of Medicine quickly emerged as forums for the input of advice to the government. President Bush moved quickly to obtain Senate confirmation for John Marburger, a physicist, to serve as his science adviser, and proceeded to fill other senior scientific vacancies in his administration.

But in the nine months since September, the complexity of the challenge that the government has set itself has only become more apparent. By declaring war on terrorism, rather than on al-Qaeda, President Bush, as he himself has grimly acknowledged, has set himself up to be held accountable if and when future terror attacks occur. Public discussion in the United States about the nature of these threats has swirled like a tornado around nuclear attacks, both fissile and 'dirty', chemical attacks and biological attacks, as well as conventional bombings, shootings and hijackings, potentially carried out by US citizens as well as by outsiders.

Limited power

In the next few days, the National Academy will issue its own recommendations of how science and technology can be harnessed to meet this cacophony of threats. It has walked some fine lines in its time, but few have been finer than the one between the need to broadcast science's relevance to the war on terrorism and the admission that science and technology have limited power to protect America.

Last October's still-unresolved anthrax attacks serve as an adequate example of what science can do and what it can't. The public confusion that accompanied the attacks showed how much government agencies need good information, as well as a scientifically competent leadership that can make statements about risk that will carry at least a modicum of public confidence.

However, the anthrax attacks themselves, which seem to have used materials from the US government's own bioweapons-research laboratories, highlighted the risks of responding to a threat by spending more money to give more people access to the knowledge and materials that constitute the threat. And the lack of any real mechanism for containing the attacks once they had occurred exposed the weaknesses of technology-led counterterrorism.

The best that science can do in most of these situations may be to provide the people who will make security decisions with timely and accurate information. This week, for example, researchers will

suggest how the United States should best use its vaccine stockpile to counter just one biological threat, that of a smallpox attack (see page 775). In the longer term, research programmes help government agencies to perform various functions, from intelligence gathering to disease inoculation, that will form part of the war against terrorism. In announcing his proposed new Department of Homeland Security on 6 June, President Bush acknowledged the importance of this role.

The details of the announcement indicate, however, that the authors of the proposal paid scant attention to the practicalities of imbuing the new agency with a sufficiently strong scientific arm. The initial proposal would build this operation mainly by transferring three existing activities — the bioterror-related activities of the National Institutes of Health (NIH), the Department of Agriculture's Animal and Plant Health Inspection Service, and the Lawrence Livermore National Laboratory — into the new department. Alarming, the administration has so far failed to make much of a case for the first two, and the third has apparently been withdrawn (see page 780).

A matter of trust

The largest proposed transfer is that of bioterror-related work from the NIH. Most of this has only recently been undertaken by the NIH — the government had doubled the size of this activity in the year before 11 September and has doubled it again since — making its transfer to the new department relatively straightforward, and even logical. However, the proposed shift raises several objections that transcend the predictable outcry that accompanies any proposal to restructure parts of the government.

The first is that the move would isolate bioterror research from existing, health-orientated research programmes at the NIH that currently house most of the relevant expertise, in everything from molecular biology to epidemiology. The second is that it would distance the research from the Public Health Service, the sister agency of the NIH whose role is regarded by most experts as central to bioterrorism containment. The third objection is that the NIH is widely trusted by the Congress and the public to spend the new money wisely — more trusted, on the basis of past experience, than a branch of a new agency such as the proposed Department of Homeland Security.

Faced with these objections, administration officials have started to backtrack from the transfer, suggesting, for example, that the NIH might continue to administer the granting of the new money under contract to the new department. But such arrangements do not work well in the US government, for a variety of reasons. The fact that the administration is resorting to them merely confirms that the research structure at the proposed new department was given little thought before the proposal was published. This impression has been confirmed by the hasty withdrawal of the massive Lawrence Livermore laboratory from the proposal.

It is disconcerting, if unsurprising, that Marburger and other senior scientific officials seem to have had no early contact with the White House cabal that put together the framework for the new department. Congress now has the opportunity to rectify the situation and construct a plan whose research component will be more deserving of the public's trust. ■





Secure sources
Radioactive materials
face tighter controls
after bomb threat
p776



Shot down
Scientists attack
plans to keep missile
test results secret
p777



Researcher quits
Controversial cat
experiments lose
group leader
p778



Star performer
Head appointed at
radio astronomy
observatory
p780

US urged to provide smallpox vaccines for emergency crews

Erika Check, Washington

The United States looks set to widen the group of health workers and other 'first responders' who will be inoculated against smallpox, as a precaution against a possible bioterror attack.

Public-health officials argue that inoculating the larger group could serve as a much-needed clinical trial of a modern, laboratory-produced smallpox vaccine, as well as strengthening the United States' ability to contain a smallpox outbreak.

A committee that advises US health officials on immunization policies may issue new guidelines on smallpox vaccination this week. The Advisory Committee on Immunization Practices (ACIP) was due to meet in Atlanta, Georgia, on 19–20 June and, according to several experts, was likely to recommend a more extensive vaccination programme. But the experts differ on which 'first responder' groups should be vaccinated, with candidates including hospital staff, police officers and emergency-response teams.

The ACIP's present recommendation is that only a small group of health workers who have to handle viruses related to smallpox be vaccinated. But many observers say that the committee is set to change its



Shot in the arm: Anthony Fauci anticipates a widening of the smallpox vaccination programme.

advice after consultation with scientists, including a public discussion organized on



Donald Henderson: the search is on for a safer vaccine.

15 June in Washington by the Institute of Medicine.

"It appears to me from town-hall-type meetings, as well as the tone of the discussion at the Institute of Medicine meeting, that we are headed towards something more than is currently stated in the ACIP recommendations," says Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases, which administers most civilian research in the United States related to diseases such as smallpox.

Fauci and others close to the immunization committee predict that it will continue to recommend against vaccinating the general public before a smallpox outbreak occurs, but that it is likely to suggest that vaccinations begin of tens of thousands of first responders who would have to investigate and contain a smallpox epidemic.

But experts differ on how widely the inoculation net should be thrown. "We have

Senate adrift on cloning as talks break down

Erika Check, Washington

Plans for the US Senate's long-awaited debate on the regulation of human cloning have collapsed, after negotiations between senators over the terms of the deliberation broke down on 13 June.

The breakdown means that prospects of any US federal legislation on cloning being passed this year are receding, with the Senate unable to muster the necessary 60–40 majority for either an outright ban on cloning or a more restricted ban, favoured by many researchers, that would permit 'therapeutic cloning' for research purposes.

The Senate's main advocate of an

outright ban, Sam Brownback (Republican, Kansas), will now try to force his fellow lawmakers to address the issue by amending unrelated bills with pieces of his legislation.

This is the strategy that he used last year to force Tom Daschle (Democrat, South Dakota), the majority leader in the Senate, to promise a vote on cloning this year. But Daschle and Brownback were unable to agree on terms for the cloning debate.

Although Senate procedures are arcane, some observers say that Brownback's new strategy is unlikely to succeed, especially as it appears to lack support from leaders of his own party. "The best way to get anything

done in the Senate is to have a structured agreement on how it will be discussed," says one lobbyist, adding that now "we have a free-for-all".

A procedural vote to kill Brownback's first attempted amendment, which would attach a ban on the patenting of cloning procedures to an antiterrorism bill, was scheduled for 18 June, as *Nature* went to press. Even if he loses it, the Kansas senator is expected to mount periodic attempts to attach a cloning ban to other bills. The House of Representatives has already backed a total ban on human cloning, which is supported by President George Bush. ■

▶ already vaccinated the groups that would go in and confirm a diagnosis, so it's a logical policy to extend that to people who are actually going to administer the vaccine if there is an outbreak," says Jonathan Tucker, director of the Chemical and Biological Weapons Non-proliferation Program at the Washington DC office of the Monterey Institute of International Studies.

The vaccine used to eradicate smallpox in the 1960s and 1970s was grown in cows, and the United States has about 77 million doses of this vaccine on hand. But 209 million doses are being prepared of a fresh vaccine that was grown in cell culture and has not been tested in large clinical trials.

"If you did it right, the trial would be large enough to learn what we need to know about the safety of the vaccine," says Anthony Robbins, chair of family medicine and community health at the Tufts University School of Medicine in Boston.

But not all public-health officials support this approach. Some argue that the risk of a smallpox attack does not justify exposing first responders to vaccine side-effects — vaccination causes serious side-effects in about 1 in 10,000 people. Others say that, given the hundreds of thousands of workers employed in health and emergency services in the United States, it is too hard to determine who should qualify for early vaccination.

"There is absolutely no end to the people who could justifiably be called first responders," says Alfred Sommer, dean of the Bloomberg School of Public Health at Johns Hopkins University in Baltimore, Maryland.

Donald Henderson, also at Johns Hopkins, a former adviser to the health secretary and leader of the global campaign that eradicated smallpox, says that the government is trying to develop a safer vaccine that could be used alone or in combination with the current vaccine. The three main candidates, he says, are the LC16m8 strain, which was given to 50,000 Japanese children in the mid-1970s; the modified vaccinia Ankara strain, which was developed in Germany in the 1960s; and a genetically engineered strain of the vaccinia virus.

After hearing the committee's advice, government officials will determine who gets the smallpox vaccine. Some politicians have already called for it to be made available to anyone who wants it. Scientists who fear the public-health consequences of such an approach will press the ACIP to issue very specific recommendations on who should receive the vaccine. ■

'Dirty bomb' scare prompts clampdown on lab security

Rex Dalton, San Diego

Universities in the United States are bracing themselves for a fresh security clampdown — on the way they handle radioactive materials used in research and medicine.

Pressure for greater security is set to grow after senior government officials said that Jose Padilla (also known as Abdullah al Muehbir), a terrorist suspect whose detention was announced on 10 June, had apparently intended to build a 'dirty bomb'. "This man thought he could get the materials from places like university labs," said Paul Wolfowitz, the deputy defence secretary, in a television interview.

New calls for tighter control of radioactive materials are now expected — hot on the heels of a law that demands tighter strictures on biological agents that could be used by terrorists. University administrators are currently analysing the Public Health Security and Bioterrorism Preparedness Act, which became law on 12 June, to assess its impact on their institutions.

"There's going to be a flurry of proposals for legislation to secure materials," predicts Michael Levi, head of strategic security at the Federation of American Scientists. "I will be surprised if there aren't much stricter standards for universities," he says. Howard Garrison, a spokesman for the Federation of American Societies for Experimental Biology, agrees that regulations may well be tightened for labs using radioactive materials.

Experts say that most laboratory radioactive material, including the liquids used in various biological and medical tests, emits only microcuries of radiation, and would be ineffective as an exploded contaminant. But many labs also house medium-level radioactive sources, emitting up to 10 curies, which could cause low-level contamination over a wide area. Some also keep strong radioactive sources that emit hundreds or even thousands of curies.



Warning sign: radioactive material is to come under closer guard across the United States.

Another possible source of highly radioactive material is the irradiator device — a heavily shielded, safe-like container containing cobalt-60 — that is used for research into food irradiation, among other purposes. But safety officers say that the theft of these machines is unlikely, as they are heavy and physically well-secured.

The Health Physics Society, a professional organization of radiation safety officers, held a discussion at its annual meeting this week in Tampa, Florida, on how best to address security problems without hampering research or health care, and has set up a task force to study the issue.

The Nuclear Regulatory Commission is responsible for licensing universities and other institutions to store radioactive materials, but it has signed agreements with 32 of the 50 US states to delegate the task to state governments. "All licensees received a written advisory after 11 September on heightened security," says John Hickey, chief of materials safety at the commission. "We are evaluating what improvements should be undertaken."

One issue that alarms Levi is the lack of government support for a project designed to retrieve highly radioactive material — much of it related to defunct nuclear research projects — from places that do not want it but cannot legally dispose of it.

In 1999, the Department of Energy (DOE) set up its Off-Site Source Recovery Project, at the Los Alamos National Laboratory in New Mexico, to carry out this task. But even though the project has identified material at more than 5,000 sites, both at universities and elsewhere, the DOE's budget proposal would cut its funding to \$2.2 million, from \$3 million this year and \$7 million last year.

The materials sought by the project are "a very likely source of radioactive material" for terrorists, Levi says. A DOE spokesman says that it will re-evaluate the project's funding in time for its next budget proposal. ■

W. HODGES/CORBIS

Geoff Brumfiel, Washington

Critics say that the Pentagon's move is an effort to deflect criticism from independent observers and Congress. They concede that the weapons system's vulnerabilities should be classified at some stage, but argue that the missile defence programme isn't close to the level of maturity to justify such classification.

Lehner says that the programme will

On 14 June, the Bush administration pulled out of the 1972 Anti-Ballistic Missile Treaty, clearing the way for a missile defence system. The next day, construction crews broke ground on the site at Fort Greeley, Alaska, that will be home to six interceptors. The Pentagon aims to provide what it terms "limited protection" from long-distance missile attack some time between 2004 and 2008. ■



From launch pad (right, inset) to impact (above) tests on interceptor missiles will soon be secret.



Declan Butler

The meeting's delegates were optimistic that the community may be ready to begin using prototypes of the databases by next January, says Catherine Cesarsky, director of the European Southern Observatory.

Enthusiasm for the IVO reflects a

The IVO will also democratize astronomy and astrophysics, adds Cesarsky, as scientists and amateur astronomers, who lack the resources to build and operate large observatories, will gain access to data from the world's best instruments and to sophisticated analysis tools.

Drug researcher quits controversial cat study

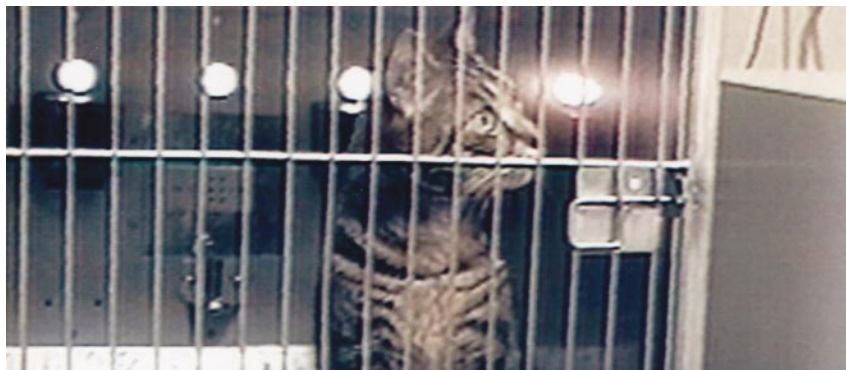
Virginia Gewin, Washington

A researcher is to turn his back on controversial research into the impact of amphetamines on cats, in the face of opposition from individuals and animal-rights groups.

Michael Podell, a veterinarian at Ohio State University (OSU) in Columbus, will no longer lead its study of the interaction between consumption of methamphetamine — better known as ‘speed’ — and the impact of feline immunodeficiency virus on cats’ brains and immune systems.

Podell’s study had been the target of lawsuits (see *Nature* **415**, 106; 2002) and, according to university officials, he has been subjected to threatening phone calls and e-mails from opponents of the work. He says that he has now accepted a “better opportunity elsewhere”, and is leaving the OSU.

Podell cites the protests as a factor in his



Animal model: researchers study cats to find a link between human drug use and HIV susceptibility.

decision, but adds that he didn’t win the backing he wanted from the OSU. “We could not come to agreement on all aspects of my career and how I should be supported,” he says.

University officials say that the OSU beefed up its security to protect Podell’s work, which is supported by a \$1.7-million grant from the National Institute on Drug Abuse (NIDA), part of the National Institutes of Health (NIH). “People are very concerned,” says Earle Holland, a spokesman for the OSU. “Nobody wanted him to leave.”

Animal-rights groups that opposed the research, including a local organization, Protect our Earth’s Treasures (POET), say they will carry on opposing the work if it continues under new leadership.

The first results from the project, published last month (*J. Neurovirol.* **8**, 240–249; 2002), show that methamphetamine drastically increases replication of the virus in cat nerve cells. Podell says this finding vindicates the approach taken by the experiment, which was designed to explore links between human drug use and susceptibility to HIV.

Supporters of biomedical research on animals fear that Podell’s departure will compromise other projects. “He isn’t the first and will not be the last” to abandon research in the face of opposition, says Frankie Trull, president of the Foundation for Biomedical Research, which lobbies against tougher regulation of animal research.

But Podell does not concede that the circumstances of his departure will encourage the harassment of researchers. “Whether I stay or go, they would continue the same tactics,” he says.

OSU officials will decide shortly whether the project, which is due to run until 2005, will continue under new leadership. Timothy Condon, associate director of the NIDA, says that the funding agency would like it to do so. “We will encourage OSU to find an alternative principal investigator for this grant,” he says.

The Physician’s Committee for Responsible Medicine, a group that supports tighter controls on animal experiments, is suing the NIH for allegedly failing to provide information about the experiment, as required by the Freedom of Information Act.

MICHAEL PODELL

Draft cow genome heads the field

David Adam, London

A biotechnology company is hoping to help farmers pick cash cows from bum steers, after rounding up the first draft genome sequence of cattle.

Researchers at MetaMorphix in Savage, Maryland, say they have not only completed a working draft of the genome, but have also identified some 600,000 single-nucleotide polymorphisms (SNPs) in the sequence.

Geneticists can use SNPs to track down links between specific genes and the traits that they control. In the cows’ case, this is likely to mean genes that make them capable of producing tender, high-quality beef.

At the moment, farmers only know after slaughter which cows produced the choicest cuts. “If we know in advance which animals are of superior quality then we could separate them out and put them on different feed regimes,” says Ed Quattlebaum, president and chief executive of MetaMorphix.

At 1X coverage, at which each base pair has been sequenced once on average, the draft is far from the finished article. A proposal to produce a publicly funded sequence of 5X to 6X coverage was recently deferred by the National Institutes of Health (NIH), but will be considered again next month (see *Nature* **417**, 473; 2002).

Alan Archibald, head of genomics and bioinformatics at the Roslin Institute near Edinburgh, UK, and part of the group behind the NIH proposal, says that the MetaMorphix announcement makes the need for a publicly accessible sequence all the more urgent.

MetaMorphix, which intends to keep its sequence and SNP data confidential, will be hunting for genes to patent, says Archibald. “At 1X coverage the sequence isn’t very impressive, but if they’ve identified 600,000 possible SNPs then that’s a very big deal indeed, and they have a significant head start,” he adds.

The SNPs still have to be verified, says Archibald, but they could allow MetaMorphix to use more powerful genetic techniques to hunt for desirable genes in normal herds of cattle. There are fewer than 2,000 SNPs in the public domain, and other researchers have to work with pedigree research herds, making it more difficult to apply the results to farm animals.

The large number of possible SNPs has taken other researchers by surprise, although many have known about the sequence for about a year. They say it was prepared by Celera AgGen, which was bought by MetaMorphix from Celera Genomics in Rockville, Maryland, earlier this year.



Have you herd? The genome sequence may help farmers locate genes that beef up their cows.

Patent flurry casts cloud over gene silencing

Erika Check, Washington

A little-known Australian company has been granted a patent that it claims will give it wide-ranging rights over the use of gene-silencing technology in mammals. And with other companies claiming rights to aspects of gene silencing — which has become a popular research tool and also holds therapeutic promise — academics fear that their access to the technology could become limited.

Gene silencing, also called RNA interference, enables researchers to shut off specific genes in cell cultures and in living organisms. Because the technique can switch off genes at different stages in development, it is a powerful tool for investigating gene function.

The company, Benitec of St Lucia in Queensland, has already been granted a patent for RNA interference in Australia. That patent is based on work done by scientists at Queensland's Department of Primary Industries which, company officials say, will ease the use of the technology in mammals, including humans.

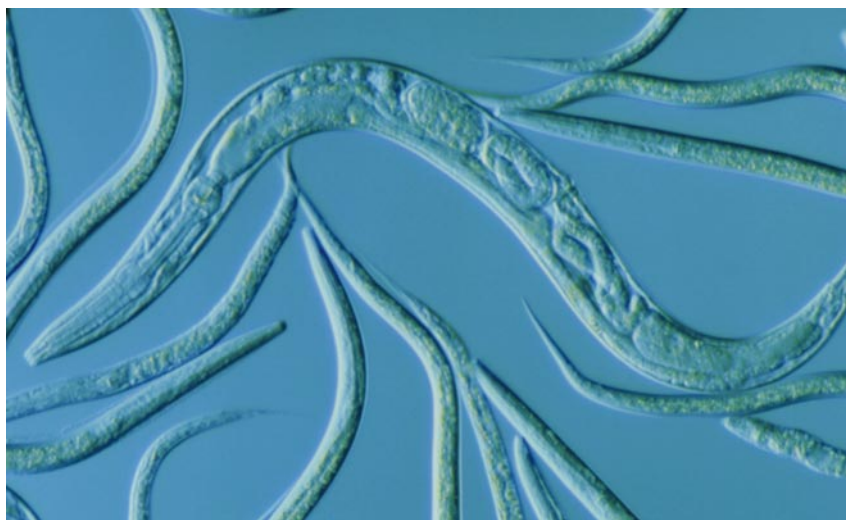
The company claims that it will soon own similar rights to the technology worldwide. But the Queensland scientists have not yet published their results, and many other RNA-interference researchers and companies say that they haven't even heard of Benitec.

Over the past few years, corporate interest in gene silencing has expanded, with a host of biotechnology start-up companies filing competing patent claims. Some scientists have eagerly joined the fray, setting up their own companies to claim patent rights. Others have stuck to their research, hoping that the patent wars will not lead to restrictions on the free use of the technology.

Most observers consider it unlikely that academics will end up having to pay to use the technology, although Benitec has not ruled out the possibility, should its claims be granted. Meanwhile, patent offices have yet to rule on earlier patent claims made by some of the field's founders.

Gene silencing is thought to be a cellular defence against viruses and 'jumping' genes that can hop into the genome, disrupting normal gene function. It is activated when the cell detects an unusual RNA, such as one with paired strands. This RNA is cut into fragments some 21 nucleotides long by an enzyme called Dicer. Single strands from these fragments, called small interfering RNAs, then bind to the alien RNA, which is disabled. But this defence mechanism can be targeted against the cell's own genes by adding double-stranded RNA with the corresponding sequences (see diagram).

It was only in 1998 that Andrew Fire, a developmental biologist at the Carnegie Institution of Washington, showed that a



Switched off: gene silencing was demonstrated in the nematode *Caenorhabditis elegans* in 1998.

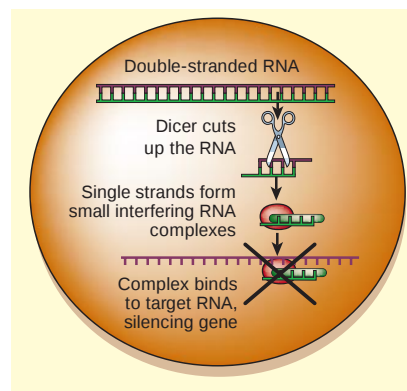
gene-silencing pathway exists in animals. Fire and his colleagues injected double-stranded RNA into cells from the nematode *Caenorhabditis elegans* and 'silenced' specific genes (A. Fire *et al. Nature* **391**, 806–811; 1998). Fire filed a US patent application in 1997, and hopes that it will be granted later this year. This application is thought to be the earliest to set out how the technology works in animals.

But Fire's technique does not work in mammals — the infusions of long RNA strands kill mammalian cells. Last year, Tom Tuschl at the Max-Planck-Institute for Biophysical Chemistry in Göttingen got around this problem by introducing small interfering RNAs directly into mouse cells (S. M. Elbashir *et al. Nature* **411**, 494–498; 2001). Tuschl's discovery attracted additional venture capital to the field, because it meant that the technology might be used to create therapies for humans.

Silent treatment

Tuschl and Phillip Sharp, his mentor at the Massachusetts Institute of Technology, applied for a patent on their RNA-interference method in 2000. Tuschl subsequently sought a patent for the use of the technique in mammals. But other groups had already applied for patents in other vertebrates, or on small pieces of RNA in invertebrates. For example, Ribopharma of Kulmbach, Germany, earned a German patent on RNA interference in April. That patent covers only short pieces of RNA.

Benitec says that it has developed a technique that entirely circumvents the problem with mammalian cells by using a vector to get long RNA strands into the cell. It uses a DNA construct that is transcribed within the cell into a single strand of RNA. This strand then



Gagging order: using double-stranded RNA, specific genes within a cell can be silenced.

folds into a hairpin shape, which is recognized by the cell as double-stranded RNA.

Benitec and Ribopharma are only the first in a line of firms getting ready to claim patent rights on some aspect of gene silencing. Another set of firms, including Nucleonics of Malvern, Pennsylvania, Devgen of Ghent, Belgium, and Cenix BioScience of Dresden, Germany, plan to profit from the technology by using it to screen for drug targets. Sharp is also said to be starting a drug-discovery company.

While these companies wait for patent offices to sort out competing claims, they are hoping to avoid large licensing fees on the technology. "The patent situation is a mess," says Tony Giordano, president of Nucleonics.

Fire is pleased that gene silencing has inspired so much excitement. But he hopes that scientists will not lose free access to it. "This is too important a technology to be restricted," Fire says. "I would be really upset if the people doing basic research had to pay a fee."

Nuclear lab relieved by Bush reversal on homeland security

Washington Just days after announcing plans to transfer the Lawrence Livermore National Laboratory to the new Department of Homeland Security, US government officials have said that it will remain within the Department of Energy.

The plans to move control of the Californian laboratory, one of the country's three nuclear-weapons research centres, were part of a reorganization of government-funded research prompted by the creation of the new agency. The decision surprised lab officials and scientists, and some feared that it could disrupt nuclear-weapons stewardship efforts (see *Nature* 417, 675; 2002).

Officials in the Bush administration, including homeland-security director Tom Ridge, put those fears to rest on 12 June, when they said that the lab will remain inside the National Nuclear Security Administration, a branch of the Department of Energy.

Antarctic researchers' voyage home is put on ice

Cape Town Dozens of Russian scientists are scanning the horizon for help after their ship became trapped in ice on its way home from their research base in Antarctica.

A rescue ship was dispatched from Cape Town on 16 June, but was expected to take at least five days to reach the area. It will then meet an Argentine naval ice-breaker, and the two vessels will have to work their way through some 250 miles of pack ice to reach the vessel, the *Magdalena Oldendorff*, and the 79 scientists and 28 crew who are stranded aboard her.

The rescue's organizers say that those on the trapped ship are in no immediate danger and have enough food and fuel for the duration of the rescue attempt, which is expected to take around a month. If the ship cannot be freed, the rescuers will attempt to



Out in the cold: scientists and crew aboard the *Magdalena Oldendorff* are awaiting rescue.

airlift its passengers, although the bitter Antarctic winter and limited daylight hours would make this difficult.

Rare atomic emission causes double delight

Paris A unusual form of radioactive decay has been observed for the first time in two separate studies in France and Germany.

Theorists predicted over 40 years ago that an atomic nucleus with an extreme ratio of protons to neutrons would decay by emitting either one or two protons. One-proton emissions were observed in the early 1980s, but scientists had not, until now, succeeded in their quest to see two-proton decays.

Using high-quality radioactive beams and sensitive detectors, researchers at the GANIL heavy-ion accelerator in Caen, France, and the GSI heavy-ion research centre in Darmstadt, Germany, say they have now observed two-proton decay from iron-45 atoms, which have 11 fewer neutrons than the most common isotope of iron. The researchers, who plan to publish their work shortly, say that their results will provide insights into the structure of atomic nuclei and help studies of astrophysical processes involving highly unstable ions.

Cash for crops will ensure that seed banks survive

Rome The genetic information contained in the world's major crop seed banks could be secured by a new fund, delegates at the World Food Summit in Rome heard last week.

In 1994, fears that crop strains that are no longer used for agriculture could disappear prompted 16 major agricultural research institutes to agree to house seed collections in perpetuity. But support for the centres has been erratic since then, so the backers of the new fund — the Food and Agriculture Organisation of the United Nations, and the Consultative Group on International Agriculture Research, an influential group of agricultural research centres — plan to raise an endowment of US\$260 million to provide permanent funding for the seed banks.

The organizations say that they are on track to raise the money by the end of the year, and that the fund could eventually be used to support smaller national seed banks, which are also under threat.

♦ www.futureharvest.org

Europe bans animal testing of cosmetics

Brussels The European Parliament last week voted to ban the testing of cosmetics on animals by the end of 2004. The directive will also ban the sale in the European Union (EU) of cosmetics tested on animals elsewhere in the world.

Before becoming law, the directive must be passed by the Council of Ministers, which represents the EU's 15 member states. Some European countries, including Britain and Germany, have already banned the use of animals in cosmetics testing. But the new directive is likely to face opposition from countries such as France, whose large cosmetics industry still tests many products on animals. It could also spark a trade dispute with countries outside the EU.

The directive calls for the use of alternative testing methods as soon as they have been validated scientifically, and increased efforts to promote and develop such methods.

Taiwan's telescope totem takes top observatory post

Washington Fred Lo, a Chinese-born astronomer who made his mark by building up Taiwan's astronomy programme, is set to lead the US National Radio Astronomy Observatory (NRAO).

Lo's appointment to the NRAO, which operates several of the world's largest facilities for ground-based radioastronomy, is expected to be announced on 20 June at a board meeting of Associated Universities Incorporated, the contractor that runs the

NRAO for the National Science Foundation.

Lo took over Taiwan's nascent Institute of Astronomy and Astrophysics in 1997 and helped to bring it to the world stage through collaboration with the Smithsonian Institution on the Submillimeter-Wave Interferometric Array. The array, a set of



Fred Lo: Taiwan's top stargazer.

eight antennae atop Mauna Kea in Hawaii, studies low-energy signals from star-forming regions. Lo will be replaced at the Taiwanese institute by Harvard University astronomer Paul Ho.

♦ www.nrao.edu

Middle East nations hope to get in synch

Paris Plans for a synchrotron facility in the Middle East appear to have survived the recent tensions in the area.

The United Nations Educational, Scientific and Cultural Organization last week announced that it was endorsing the creation of SESAME, the Synchrotron-light for Experimental Science and Applications in the Middle East, which is to be based in Allaan, near Amman in Jordan. The machine will be an upgraded version of the BESSY I synchrotron, which has been donated by Germany.

The centre will become a legally

independent lab when six of the project partners, which include Iran, Israel and Egypt, accept the terms for its use. "As Arabs we realize that acquiring technology is very important for the scientific and economic advancement of our people," says Said Assaf, head of the Arafat National Scientific Center for Applied Research in Ramallah. "It is also a bridge to peace through science, and working jointly with Israeli scientists on SESAME will help to advance this."

♦ www.sesame.org.jo

Flurry of new planets has astronomers feeling jovial

Washington The prolific planet-hunting team of Geoff Marcy and Paul Butler last week announced the discovery of 15 planets in foreign solar systems — including one with properties similar to those of Jupiter.

Marcy, of the University of California, Berkeley, and Butler, of the Carnegie Institution of Washington, found an object of three to five times Jupiter's mass orbiting the star 55 Cancri, 41 light-years from Earth.

What's more, it orbits its star at roughly the same distance at which Jupiter orbits our own Sun. The team observed 55 Cancri for 13 years to establish the planet's orbit, by far the longest they have spent observing a star.

Around 90 extrasolar planets have now been identified, two-thirds of them by Marcy and Butler's team. Also among the new finds is the smallest planet yet discovered outside our Solar System. It orbits the star HD49674 in the constellation Auriga and has a mass that is roughly 40 times that of Earth.

♦ <http://exoplanets.org>

Murder? Call the crime-busting botanist

San Diego A chance answering of a telephone call turned a Californian botanist into a detective for a day, and helped police to find a the body of a murdered child.

Police were searching for a missing 11-year-old girl, whose father killed himself on 29 May in his truck in the Sierra Nevada Mountains, north of Sacramento. The only clue was some plant debris found in the truck. Botanist Kristina Schierenbeck, who answered the call at California State University at Chico, examined the debris and determined that the plants came from an area at an altitude of 800–1,200 metres, on a west-facing mountain slope on which coniferous trees and chaparral merge, near to a water source.

Schierenbeck, who had never helped the police before, drove through the mountains with a detective looking for the right plant combinations. At their fifth stop, the botanist walked 20 metres from the car and found the body. Schierenbeck says she is glad she could help, but hopes that she isn't called again.

Squaring up over ancient life

The textbooks say that oxygen-producing microorganisms evolved some 3.5 billion years ago. But as that claim and its author come under attack, the history of life on Earth may have to be rewritten. Rex Dalton investigates.

It was the academic equivalent of a heavy-weight prizefight. In the red corner, defending his title as discoverer of the Earth's oldest fossils, was Bill Schopf of the University of California, Los Angeles (UCLA). In the blue corner, Martin Brasier of the University of Oxford, UK, who contends that Schopf's 'microfossils' are merely carbonaceous blobs, probably formed by the action of scalding water on minerals in the surrounding sediments.

The bout took place on 9 April at the second Astrobiology Science Conference, held at NASA's Ames Research Center in Moffett Field, California. It was the first time that Schopf and Brasier had faced each other since their publication, in the 7 March issue of *Nature*, of conflicting papers^{1,2} re-examining Schopf's claim to have discovered cyanobacteria in Australian rocks 3.5 billion years old.

Most judges gave a clear points victory to Brasier. But as Schopf licks his wounds, one of his former graduate students has come out of the woodwork to deliver what may prove to be a knockout blow. She argues that her supervisor was aware from the start of evidence that cast doubt on his conclusions — a charge Schopf vigorously denies.

The samples in question are Schopf's claim to scientific fame — they are even listed in *Guinness World Records* as the world's oldest



Trading blows: Martin Brasier (left) puts his case at the NASA meeting as Bill Schopf (right) listens.

est fossils. But more significant than their sheer antiquity is Schopf's interpretation that they are most likely to be fossilized cyanobacteria — organisms that can use the energy of sunlight to make sugars from carbon dioxide. Although other researchers have found microorganisms in rocks of similar age^{3,4}, these are not thought to have been capable of this trick. The existence of cyanobacteria 3.5 billion years ago, which would have pumped oxygen into the atmosphere as a by-product of photosynthesis, has long been a puzzle — the geochemical evidence suggests that the atmosphere contained very low levels of oxygen until about a billion years later^{5,6}.

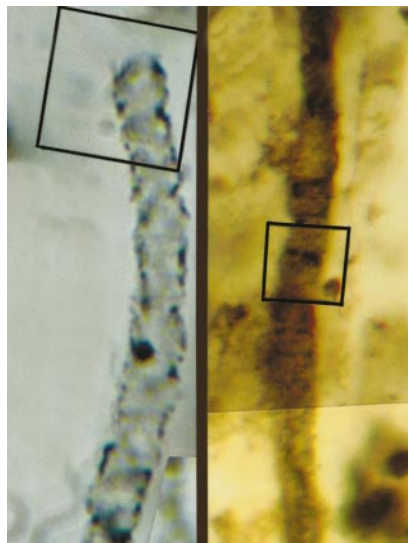
Slice of life

The specimens described by Schopf were collected from chert — a flint-like sedimentary rock — near the Western Australian town of Marble Bar (see map, opposite). Together with sites in Greenland and South Africa, this region boasts the world's oldest surface rocks.

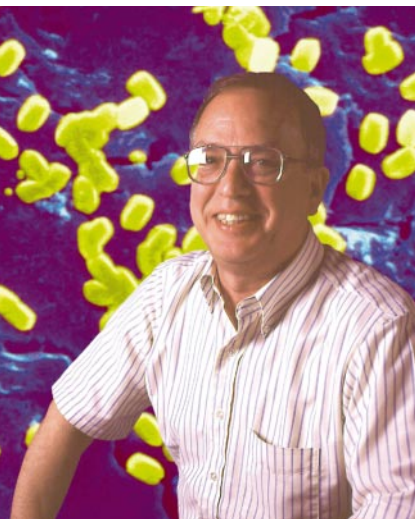
Looking for fossil microbes in such samples involves cutting slices thin enough to shine a light through, and then examining them under the microscope for countless hours, micrometre by micrometre. Schopf and his graduate student Bonnie Packer first suggested that the samples may contain cyanobacteria in 1987 (ref. 7). In 1992, Schopf

advanced his description⁸ in a book, *The Proterozoic Biosphere*, which he co-edited. Then in a *Science* paper⁹ the following year, Schopf described a total of 11 taxa of microorganisms from the samples, arguing that most were filamentous colonies of "probable cyanobacteria" dating from 3.465 billion years ago.

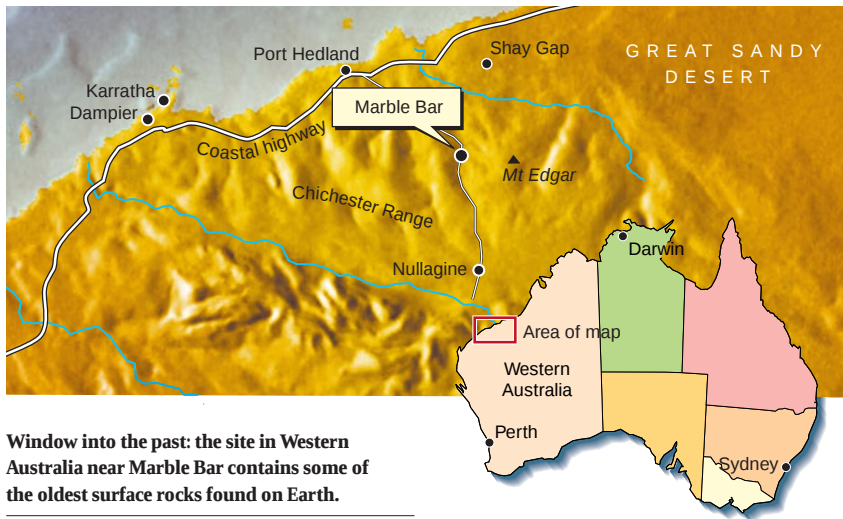
At the time, some researchers thought it



Still life: photomicrographs claimed by Schopf to show 3.5-billion-year-old microbial fossils.



Eye of the storm: Bill Schopf claims to have discovered the oldest fossils of life on Earth.



Window into the past: the site in Western Australia near Marble Bar contains some of the oldest surface rocks found on Earth.

was a stretch to conclude that the specimens were cyanobacteria based on morphology alone. But over the years, Schopf's description became textbook orthodoxy. However, when Brasier updated his microfossil textbook¹⁰ in 1999, he re-examined Schopf's specimens, which are now stored at the Natural History Museum in London (see 'The London connection', overleaf). As Brasier searched through the thin sections, he found that some of the 'fossils' were branched in ways never depicted in Schopf's papers. Others took on weird shapes that looked quite unlike filamentous cyanobacteria. "You scratch your head and wonder if maybe there was some terrible mistake," says Brasier, recalling his thoughts at the time.

His suspicions raised, Brasier pored over the specimens, giving some nicknames such as 'Loch Ness monster' and 'wrong trousers'. What he saw under the microscope was buttressed by a 1999 trip to Marble Bar, which revealed that the site from which the specimens came was riven with lava flows, rather than being a continuous sedimentary formation.

From their geological mapping and chemical analyses, Brasier and his colleagues concluded that Schopf's samples had not

originally been laid down on the floor of a shallow sea, as his papers suggested. Instead, the site seemed to have been a hot spring pouring volcanically heated water from the seabed. Schopf's 'microfossils', Brasier argued, were merely artefacts formed from amorphous graphite in this hostile environment.

Rocks revisited

Having submitted his paper² to *Nature* in February 2001, Brasier began lecturing about his findings on the conference circuit, including a talk at the Earth System Processes conference in Edinburgh in June 2001, organized by the Geological Society of London and the Geological Society of America.

Early last year, as he became aware of Brasier's results, Schopf began his own reanalysis. He had earlier teamed up with researchers from the University of Alabama at Birmingham who were experimenting with laser-Raman imagery. This is a technique for characterizing molecular structure, in which a laser is focused on a point within the sample and optical sensors record the spectrum of the backscattered light. In January 2001, Schopf and his Alabama colleagues reported that these spectra can distinguish between microfossils and inor-

ganic carbonaceous deposits¹¹. They then applied the same technique to Schopf's 'microfossil' samples, reporting in their *Nature* paper¹ that the results substantiated the original claim that the specimens were fossilized microorganisms.

Given that the search for extraterrestrial microbial life is likely to involve similar techniques to those used to look for microfossils, it was natural that NASA's April astrobiology conference would be the venue for Schopf and Brasier's first public duel since their papers were published. In front of a packed audience in a huge white tent, the pair gave back-to-back 15-minute lectures. It amounted to an intellectual brawl, with the protagonists exchanging several low blows.

Schopf hit out at Brasier for "errors" in interpretation — the specimens were not branched, but "folded". Brasier, every inch the haughty Oxford don, was brutally dismissive of his opponent's arguments: "A truly hydrothermal performance, with more heat than light," he announced, after taking over from Schopf at the podium.

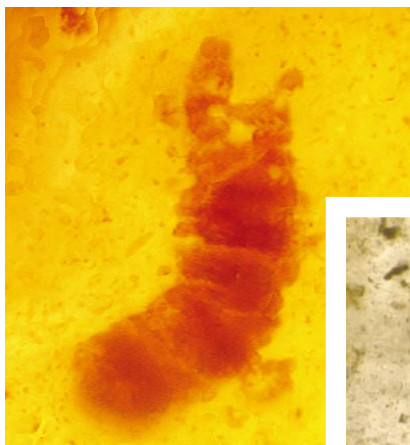
Seeking to convince the gathering of some 300 scientists that the specimens were, indeed, microfossils, Schopf had adopted the tone of a revivalist preacher, shouting at times. "Hallelujah! I believe!" responded one palaeontologist mockingly, after Schopf brought his talk to a close. But when both men had finished their presentations, it was clear that Schopf had won few converts to his cause.

Elements of doubt

Those present also thought Schopf had made a major concession, accepting that the specimens were not oxygen-producing cyanobacteria after all. In his talk, he stressed that none of his papers claim that the specimens are definitely cyanobacteria. "I was absolutely struck by that," says geochemist John Hayes of the Woods Hole Oceanographic Institution in Massachusetts. The original suggestion that the specimens were cyanobacteria was the key point of interest, and if that is gone, many in the field are unmoved by the issue of whether they are fossils or not. "That makes his whole argument irrelevant," says Australian geologist Roger Buick, now at the University of Washington in Seattle.

Other presentations in the same session also had a bearing on the issue of Schopf's specimens. Physicist Stephen Hyde of the Australian National University in Canberra showed that, using techniques pioneered by Juan Garcia-Ruiz of the University of Granada in Spain, his team has created at room temperature mineral deposits that under the microscope look exactly like Schopf's purported microfossils. "That was remarkable material," says Hayes.

And although the laser-Raman evidence was not seriously challenged at the Ames meeting, experts in the field are now questioning Schopf's claim that the technique



Mistaken identity? Brasier believes Schopf's 'fossils' are merely artefacts caused when the rocks formed.

The London connection

As micropalaeontologists debate whether Bill Schopf's Australian specimens really are fossils, some experts are also questioning why they haven't been returned to their country of origin. After studying the thin slices of rock for several years, Schopf deposited them in 1992 at the Natural History Museum in London (below). In 1988, Australian law was changed to require the return of fossils collected from its territory and used to describe new species. It does not apply to specimens collected before this date, but Bruce Runnegar, an Australian who now directs the Center for Astrobiology at the University of California, Los Angeles, says he cautioned Schopf against sending the samples to London: "I advised him not to do it."

The Geological Survey of Western Australia in Perth wrote to Schopf seeking the return of the specimens. And in 1999, when Schopf published an acclaimed account of his work¹⁶, he wrote that "backup specimens" had been sent to Australia. But Kath Grey, the survey's palaeontologist, says that the specimens sent to Perth appear to be rejects in which no identifiable 'microfossils' have yet been found.

A spokeswoman for the Natural History Museum says that officials there would be glad to discuss the issue of repatriating the samples. But Grey remains concerned about the condition of the much-travelled specimens. In an unusual move, the London museum last year permitted all of the specimens — not the normal allotment of three — to be shipped to Schopf for fresh analysis.

Two of the thin sections were broken while on loan or in transit. "I'm amazed they sent them to Los Angeles," says Grey. "We wouldn't have allowed typed material to have been sent overseas. We like people to come to the museum; it lessens the risk of breaking."



can provide a signature of once-living material. "I don't know of any work that definitively shows the difference between the Raman spectra of organically precipitated material and inorganically precipitated carbon," says Jill Pasteris, a geologist at Washington University in St Louis, Missouri, who has worked on laser-Raman for 15 years. "Laser-Raman is a wonderful technique, but they pushed it too far. Schopf's *Nature* article is a misstatement of what the technique can do." Pasteris also notes that Brasier's own laser-Raman analysis² found no differences between the spectra of the purported fossils and those of neighbouring inclusions in the rocks.

But Schopf, responding to *Nature*'s written questions, stands by his paper. "The claim of our article is that morphology and chemistry together provide a powerful means to address the problem of biogenicity," he wrote.

As the dust settled after the meeting session, Brasier remained aghast at Schopf's claim that the fossils were folded, rather than branched. If so, why had Schopf not pointed this out in his earlier publications? "He was arguably misleading the scientific community and the public about their true nature," says Brasier. Schopf rejects that accusation, adding that the folding conveys "no useful scientific information".

But now Packer, who has not previously spoken out, claims that Schopf was, indeed, highly selective with the evidence he presented in his original papers. Backing up her account with pages from her lab notes, Packer claims that Schopf withheld from publication images of the purported cyanobacteria that indicated branching structures.

Although named as co-author on the initial paper¹, Packer says that she became increasingly concerned about Schopf's presentation of the data. She claims that her attempts to challenge him met with stubborn resistance. "There wasn't a bloody thing I could do," she says. Packer parted company with Schopf in 1987, and eventually completed her PhD in another UCLA laboratory. She now works at the US Army Environmental Center at the Aberdeen Proving Ground in Maryland, studying the environmental impacts of unexploded and detonated ordnance.

Photographic memories

Again responding to written questions, Schopf initially contested Packer's account of events: "Never, not at any time, did she mention to me (or show me photos of) branched filaments ... Dr Packer's memory is simply in error." But shortly before this article went to press, he sent a second e-mail, saying that he had found records of Packer's observations of "branching". In this note, Schopf added that he could not recall whether he saw the photos at the time, which he argued depict "mineralic fossil-like artifacts" or "clumps of unbranched overlapping poorly preserved filaments". But Packer says that Schopf's handwriting is on the photos she retains.

Supporters and critics of Schopf alike describe him as a driven and tenacious character — nicknamed 'Bull' Schopf by some — whose energy and enthusiasm has done much to raise the profile of micropalaeontology, and to draw funding into the field. "He has a driving ambition to be in the limelight, and he doesn't like to admit he's wrong," says

one former colleague. But these traits have led Schopf into conflict with his collaborators on at least one previous occasion.

In 1980, Schopf was working with Stan Awramik, now an associate vice-chancellor at the University of California, Santa Barbara; Malcolm Walter, now director of the Australian Centre for Astrobiology at Macquarie University in Sydney; and Buick, then a graduate student at the University of Western Australia in Perth. Schopf prepared a manuscript claiming the discovery of 3.5-billion-year-old bacterial fossils, after analysing rocks collected in 1977 by Awramik in the same general region as the specimens in the Brasier dispute. They came from sites, one of which had been identified by Buick, near a structure called the North Pole Dome, named for its desolate appearance. Buick says he returned from the field one day to learn that the manuscript had been accepted for publication by *Science* — and that he was listed as an author. "I wrote to *Science* saying I strongly disagreed, and withdrew my

name as co-author," Buick says.

Science subsequently declined to publish the paper. Undeterred, Schopf, Awramik and Walter eventually published their findings in *Precambrian Research*¹². But before the paper appeared, Awramik and Schopf had become embroiled in a dispute over credit for the specimens' discovery. Although Awramik was, in the end, satisfied with the credit given for his contribution, he never collaborated with Schopf again. Buick, meanwhile, published a rebuttal to the paper a few months later¹³, and the debate continued in print for some time^{14,15}.

Although that disagreement has since subsided, Schopf and Brasier are not calling it quits, and are set for a rematch. Their next bout is scheduled for the International Conference on the Origin of Life, to be held in Oaxaca, Mexico, from 30 June to 5 July. Don't expect either man to pull his punches. ■

Rex Dalton is *Nature*'s US West Coast correspondent.

- Schopf, J. W., Kudryavtsev, A. B., Agresti, D. G., Wdowiak, T. J. & Czaja, A. D. *Nature* **416**, 73–76 (2002).
- Brasier, M. D. et al. *Nature* **416**, 76–81 (2002).
- Knoll, A. H. & Barghoorn, E. S. *Science* **198**, 396–398 (1977).
- Walsh, M. M. *Precamb. Res.* **54**, 271–293 (1992).
- Copley, J. *Nature* **410**, 862–864 (2001).
- Hayes, J. M. *Nature* **417**, 127–128 (2002).
- Schopf, J. W. & Packer, B. M. *Science* **237**, 70–73 (1987).
- Schopf, J. W. in *The Proterozoic Biosphere: A Multidisciplinary Study* (eds Schopf, J. W. & Klein, C.) 25–39 (Cambridge Univ. Press, New York, 1992).
- Schopf, J. W. *Science* **260**, 640–646 (1993).
- Brasier, M. D. *Microfossils* (George Allen & Unwin, London, 1980).
- Kudryavtsev, A. B., Schopf, J. W., Agresti, D. G. & Wdowiak, T. J. *Proc. Natl Acad. Sci. USA* **98**, 823–826 (2001).
- Awramik, S. M., Schopf, J. W. & Walter, M. R. *Precamb. Res.* **20**, 357–374 (1983).
- Buick, R. *Precamb. Res.* **24**, 157–172 (1984).
- Awramik, S. M., Schopf, J. W. & Walter, M. R. *Precamb. Res.* **39**, 303–309 (1988).
- Buick, R. *Precamb. Res.* **39**, 311–317 (1988).
- Schopf, J. W. *Cradle of Life: The Discovery of Earth's Earliest Fossils* (Princeton Univ. Press, Princeton, New Jersey, 1999).



Full house

Across the world, animal facilities are overflowing with mutant mice. Jonathan Knight and Alison Abbott consider a logistical nightmare that is reaching crisis point, thanks to the revolution in genomics.

The new mouse facility at Baylor College of Medicine in Houston, Texas, is state of the art. It has robotic cage washers and separate rooms for experimental procedures. Opened in September 2000, it holds 40,000 cages. But by next spring, all those cages will be full.

It's the same story at research centres all over the world, with mice filling up the space as fast as the bricks are laid. The GBF, the National Centre for Biotechnological Research in Braunschweig, Germany, opened a new facility for 8,000 mice less than two years ago, and it's already full. A second mouse house, which will nearly double capacity, is scheduled to open in October. "And that will be full by the end of next year," says Werner Müller, who runs the centre's mouse facilities.

Overworked animal technicians can blame genomics. On 6 May, an international sequencing consortium posted 95% of the mouse genome sequence on the Internet. And on 31 May, Celera Genomics of Rockville, Maryland, published a paper on one chromosome of its rival mouse genome sequence (R. J. Mural *et al. Science* **296**, 1661–1671; 2002). But turning raw sequence data into functional information means creating many thousands of mutant mice in which individual genes are disabled. And mice take up a lot more space than fruitflies or nematode worms — a single shoebox-sized cage holds only four or five animals, and has to have space around it for ventilation and easy access by handlers.

The number of labs making and studying knockout mice — in which a particular gene is targeted and disabled — has risen steadily since the technique was developed in the late 1980s. Around 3,000 knockout strains are now available, and the number is growing exponentially.

Yet this flood of new mutants is barely a trickle compared with the inundation that is to come. As well as creating knockouts, researchers can feed the chemical *N*-ethyl-*N*-nitrosourea (ENU) to male mice to cause small random mutations in their sperm. These often don't completely disable the gene but change the properties of the protein it encodes. The treated males are bred and their offspring, each of which could carry a unique mutation, are screened for interesting characteristics, or phenotypes. Very large

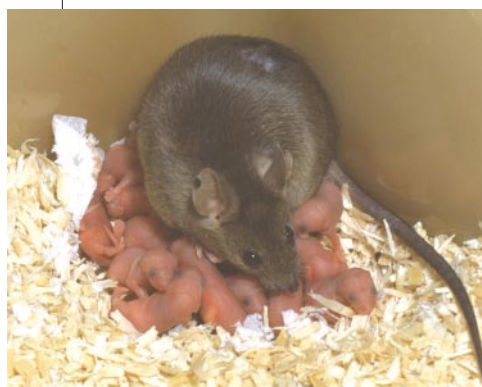
mutagenesis screens are in progress on four continents. The first, and still the biggest, is in Munich, where 28,000 mutants have been screened since 1998 for phenotypes ranging from allergy to changes in behaviour.

No limits

Geneticists agree that the two approaches are complementary. Whereas knockouts are the faster way to find out what a given gene might do, random mutagenesis is more efficient at finding genes involved in a particular phenotype. Mutagenesis screens can also turn up genes that a computer algorithm designed to scan the genome for sequences that look like genes would miss. And mutagenesis can generate different small mutations in the same gene, which enables researchers to investigate subtler phenotypes than those produced by the sledgehammer knockout. Mouse geneticists may eventually get what they've always wanted — one or more mutants for every gene in the genome.

Supposing an average of five useful mutations per gene, and assuming that the mouse has 30,000 genes, that would mean 150,000 distinct genotypes. Pierre Chambon, director of the Institute for Genetics and Molecular and Cellular Biology in Strasbourg, France, estimates that breeding and working on this many strains of mice would need at least 60 million animals.

The logistics become even worse, because the effects of a knockout or mutation may need to be studied in different mouse strains. Mouse geneticists also use extra tricks to turn



Population explosion: each mouse strain needs several hundred individuals to keep it going.

their mutants into better models of human disease. Since the 1980s, they have been making transgenic mice by inserting foreign genes into fertilized mouse eggs. This approach is particularly useful when combined with knockouts; one can see if a human version of the gene can substitute for the knocked-out mouse version. And, since the early 1990s, 'conditional' knockouts have let researchers disable a target gene in a particular tissue at any time in the mouse's life. "There is really no limit to the number of strains that could be generated — you can expand to infinity," says Chambon.

Except that the space available is far from infinite. In practice, only a fraction of mutants would be under investigation at any time; the rest would be in storage as frozen sperm or embryos. But university mouse houses and central storage facilities are already creaking under the strain. "There is a chronic shortage of storage for mutants," says Steve Brown, director of the UK Medical Research Council's Mammalian Genetics Unit at Harwell, near Oxford, which houses the world's second-largest ENU screen. "It is a worldwide problem — we'll soon be swamped and forced to lose mutants, which is a pity."

This already seems to be happening. The Jackson Laboratory in Bar Harbor, Maine, which serves as the principal repository for mutant mice in the United States, is accepting a smaller percentage of applications for storage each year. Currently, Jax — as the lab is



Steve Brown (left) fears losing mutant strains, and Barbara Knowles at Jax (below) is already turning away storage applications.

known — can accept about 70 new knockout strains a year, says Barbara Knowles, Jax's director of research. Last year, the lab turned down 50 applications. Although most of these were rejected for scientific reasons, such as duplication of an existing strain or unpublished supporting data, a few couldn't be accommodated simply because of lack of space and funding. Tak Mak, who studies knockout mice at the University of Toronto's Ontario Cancer Institute, says that he stopped trying to store animals at Jax years ago. "I got rejected too many times."

Deep freeze

As a result, academic researchers are having to store their mutant mice on campus. One consequence of this is that mouse strains tend to be lost when a project ends or a postdoc or graduate student moves on. "We've stopped keeping old mice around," says Richard Flavell, an immunologist at Yale University in New Haven, Connecticut. He says it's not unusual to receive a request for a mouse mutant that is mentioned in a publication but is no longer being kept.

So what are the solutions? Greater use of cryopreservation is one possibility. Frozen embryos take up very little space. But to get the 500 embryos needed to preserve a strain reliably means mating around 200 females, limiting the number of strains that can be archived each year. Jax now keeps most of its strains frozen, transferring embryos to surrogate mothers when it receives an order, which means that biologists have to wait up to six months to get their mice.

Sperm can also be frozen and, in theory, just a few males can produce enough sperm to preserve the strain. But freezing and thawing weakens sperm, making *in vitro* fertilization less successful. Thawed sperm have trouble penetrating the egg's protective coat, the zona pellucida. Possible solutions include injecting the sperm directly into the egg or weakening the zona with a laser beam or solubilizing chemicals. But there is still a way to go. "It's still not as efficient as we would like," says Larry Mobraaten of Jax.

With no great scientific advances on the horizon, researchers argue that funding agencies will have to spend more money on expanding and upgrading animal facilities. Adriano Aguzzi, who works on mouse models of prion disease at the University of Zurich in Switzerland, says that his Institute of Neuropathology spends more than 400,000 euros (US\$377,000) a year maintaining its mouse colony. The university pays half, but the rest must be found from agencies that seem reluctant to invest in infrastructure. "The problem is getting bigger every day," says Aguzzi. "The pressure on space and money is a nightmare."

The US National Institutes of Health (NIH) is now responding to the impending crisis. In May, it agreed to pay for the storage



Fertile ground: Larry Mobraaten is trying to boost the effectiveness of frozen sperm for IVF.

and distribution of the mutants produced by an ENU screen for neurological mutants coordinated from Northwestern University in Evanston, Illinois. The size of the grant is still under negotiation, but it will include money for a centralized ordering facility and for cryogenic storage at Northwestern, Jax and the Oak Ridge National Laboratory in Tennessee.

The NIH is also providing more than \$3 million per year to sponsor the Mouse Mutant Regional Resource Centers, a collection of small academic and privately operated facilities for storing and distributing mutants. Meanwhile, the European Mouse Mutant Archive, a distributed network of centres with its hub at Monterotondo, near Rome, is on its way to becoming the European Jax.

But in the long run, the pressure may only be eased by technologies that silence genes at will, allowing functional studies to be carried out without the need for thousands of mutant strains. For example, RNA interference, a technique that silences genes by using short stretches of double-stranded RNA to block their messenger RNAs, was initially thought not to function in mammals. But recent developments have raised hopes of making it work reliably in mice.

Other researchers are banking on the identification of small molecules that selectively disable the protein produced by the gene of interest. The humble mouse, it seems, is already being lined up for a key role in biology's next revolution — that of proteomics. ■

Jonathan Knight is Nature's contributing correspondent in San Francisco; Alison Abbott is Nature's senior European correspondent.

Jackson Laboratory

♦ www.jax.org

European Mouse Mutant Archive

♦ www.emma.rm.cnr.it



Learning from past mistakes about nuclear waste

Any managing organization must be seen to be independent, transparent and legitimate.

Sir — Your News in Brief “Britain urged to dump nuclear-waste agencies” (*Nature* **417**, 111; 2002), reporting the Royal Society’s call for reform of the institutions overseeing nuclear waste to restore public confidence in them, is consistent with the position of Nirex (the UK radioactive waste management agency), that the government should set up a new radioactive waste management organization independent from the nuclear industry. This is contained within the Nirex submission to the UK Department for Environment, Food and Rural Affairs consultation, which can be found on our website at www.nirex.co.uk.

Nirex has undergone its own reform to learn the lessons from past failures and ensure that, in future, whatever organization is responsible for managing nuclear waste in the United Kingdom, it is independent, transparent and above all legitimate in the eyes of the public.

Nirex was last in the public eye during the planning proposal in 1997 to build a ‘rock-characterization facility’ to examine suitability of the Sellafield site in northwest England for an underground nuclear-waste repository. The whole process was a resounding failure, for which Nirex accepts a significant part of the blame.

Since then, Nirex has done much work

to try to learn from this: establishing an independent transparency panel; introducing a whistleblowers’ policy; and holding regular dialogue with environmental groups. Stakeholders have generally welcomed these initiatives, which need to be built on in the future. An independent report by the consultancy Environmental Resources Management on stakeholder attitudes is on the Nirex website.

If people wish to find out more or become more actively involved, I would urge them to contact me at Nirex.

Chris Murray

Nirex, Curie Avenue, Harwell,
Oxfordshire OX11 0RH, UK

To sell science, find out what people want to buy

Sir — Elizabeth Tait in Correspondence (*Nature* **417**, 221; 2002) discusses ways in which scientists and science communicators can work together. I would like to add that the current political pressure on scientists to become more market-oriented is making them reposition themselves and learn about research marketing. To this end, science communicators are supposed to make the public aware of the value that basic science contributes to society.

But are these activities ‘research marketing’? In his famous essay “Marketing myopia” (*Harvard Business Review* **38**, 465–56; 1960), Theodore Levitt argued that an industry’s primary aim should be to satisfy customers, not to sell goods. This proposition should hold especially true for a service industry — and basic science is ultimately no more than that.

Scientists can learn from Levitt to serve the needs of their customers, the public. According to Levitt, the difference between selling and marketing is that the former fulfils the needs of the seller and the latter the needs of the buyer. Academic scientists must stop trying to sell their achievements to the public, and instead market them by considering public needs right at the beginning of their endeavours.

However, scientists work on the not-quite-yet-possible, whereas public discussions of ethics and politics take place only when scientific results are just becoming possible. Hence there will be always a gap between scientific advances and societies’ judgements about them. If

scientists want to contribute to society, and also see the application of the knowledge they generate, they need to be less concerned with the ‘public understanding of science’ and more with scientists’ understanding of the public.

To this end, market forces can contribute significantly to the understanding of what the public expects and desires from science. Therefore, scientists conducting basic science should expand their agreements, collaborations and partnerships with industrial sponsors and learn from them. Commercial companies can provide academic scientists not just with money, but with knowledge of what the market or, better, the public wants. Market forces will help to establish how much the public is willing to pay for advances in knowledge and technology.

Till C. Jelitto

PR&D — Public Relations for Research and Development, Margaretenstr. 70, A1050, Vienna, Austria

Macroecology: new, or biogeography revisited?

Sir — According to his News and Views report¹ of a recent symposium, Sean Nee and the British Ecological Society believe that macroecology is an infant discipline, a child of the conjunction of ecology and evolution. Yet the scope of macroecology is largely contained within the relatively old discipline of biogeography, which dates back at least to Alexander von Humboldt in the early nineteenth century. Some of the symposium’s participants have written books about

biogeography (see, for example, refs 2, 3). As Nee states, the difference between ecology and macroecology is largely a matter of scale, both in time and space. The bigger picture is, effectively, biogeography.

Howard J. Fisher

Department of Geography, Avondale College,
Cooranbong, New South Wales 2265, Australia

1. Nee, S. *Nature* **417**, 229–230 (2002).
2. Brown, J. H. and Lomolino, M. V. *Biogeography* 2nd ed. (Sinauer, Sunderland, Massachusetts, 1998).
3. Whittaker, R. J. *Island Biogeography* (Oxford Univ. Press, Oxford, 1998).

Online database could end taxonomic anarchy

Sir — In his Commentary “Challenges for taxonomy”, H. C. J. Godfray¹ presents compelling arguments for centralizing and revitalizing taxonomy using an Internet database of all taxon names and their descriptions, integrated with other biologically relevant information. He reasonably proposes that such a mammoth project would need to be undertaken in manageable stages, for example one group at a time. However, he does not suggest any logical starting points.

One obvious group is bacteria. The Bacteriological Code² already centralizes bacterial taxonomy by ensuring that all name changes are either published in the *International Journal of Systematic and Evolutionary Microbiology* (*IJSEM*) or, if published elsewhere, are validated by announcement in that journal. Hence Godfray’s innovative vision of what could be called a ‘species bank’ could be realized quite easily for bacteria by

compiling a database of all current bacterial names, and integrating it with an online version of *IJSEM* to incorporate ongoing revisions. Given the importance of bacterial taxonomy in fields within and outside biology, such a database would be extensively accessed by 'pure' science, medicine and industry, and might thus be able to attract broad logistical support. Life began as bacteria, and it might be fitting if our comprehensive inventory of life also began with this group.

Although Godfray is correct to say that that current taxonomic codes prohibit "purely electronic description", the current version of the Zoological Code officially recognizes new names posted on websites as long as five hard copies are deposited in libraries³. However, such technophilia has permitted taxonomic anarchy. The ease of electronic publishing has encouraged some individuals to name electronically a plethora of dubious new species, in groups on which they have little taxonomic expertise (I do not cite these websites here, to avoid drawing further attention to them). The resultant mess will take decades to clear up. A central official taxonomic database, peer-reviewed and managed along the lines suggested by Godfray or by other means to this end, would solve this problem, facilitating rapid information dissemination via the Internet while at the same time filtering out unscrupulous taxonomic practices.

Michael S. Y. Lee

Department of Environmental Biology,
University of Adelaide, North Terrace,
Adelaide 5000, Australia

- Godfray, H. C. J. *Nature* **417**, 17–19 (2002).
- Sneath, P. H. A. *International Code of Nomenclature of Bacteria* (International Union of Microbiological Sciences, Washington DC, 1992).
- International Commission on Zoological Nomenclature. *International Code of Zoological Nomenclature*, 4th edition (The International Trust for Zoological Nomenclature, London, 1999).

Science's policy on access to private data

Sir — In their thoughtful and welcome Commentary on data access "The times they are a-changin'" (*Nature* **417**, 589–590; 2002), Ari Patrinos and Dan Drell make two statements about *Science's* publication policy that require correction.

With respect to the arrangement that allowed Celera Genomics to provide its sequence data through its own website, Patrinos and Drell say that the company limited free access to one million base pairs

per day. In fact, Celera announced on its website that it would provide, and it did provide, compact discs containing the entire sequence to those who asked.

Patrinos and Drell also say that "those in the private sector have to negotiate a fee as a prerequisite for access". That is simply wrong — commercial accessors could and did receive the sequence free on executing a material transfer agreement that committed them not to distribute it or to use it for new research projects. That arrangement, incidentally, followed criticisms from members of the publicly funded Human Genome Project that any arrangement that required commercial accessors to pay would constitute "discriminatory licensing". The agreement we at *Science* reached with Celera circumvented that objection, thus meeting the challenge aptly described by Patrinos and Drell: "to suggest how the private sector can be persuaded to share more data, to the benefit of all".

Donald Kennedy

Editor-in-Chief, *Science*, 1200 New York Avenue
NW, Washington DC 20005, USA.

Curiosity and generosity of a great scientist

Sir — I read with sadness about the passing of Victor Weisskopf (*Nature* **417**, 396; 2002). Twelve years ago, in San Juan, Puerto Rico, I was sitting next to an empty seat on a plane, wondering which of the many oversized people with too much carry-on baggage I would have to put up with for the flight to Boston. An elderly man sat down next to me and we began chatting: it was Weisskopf.

For five glorious hours I — who as a teenager had wanted to be an astronomer before I realized that my mathematical abilities were woefully inadequate to the task — plied Weisskopf with questions about physics and cosmology. In his turn, he posed penetrating questions about my own field of large-whale biology.

Weisskopf had all the hallmarks of a great scientist. He was ungrudgingly generous with his own vast knowledge and happy to share his insights with an interested stranger; and he was insatiably curious about everything in the world around him, whether or not relevant to his own work.

The unique and fortunate nature of my experience that day was not lost on me. After all, how often does one get the opportunity to ask one's seatmate, "So, do you think that general relativity and quantum theory will be unified in the foreseeable future?" and have the

expectation of an informed reply?

The answer, by the way, was "No".

Phillip J. Clapham

Large Whale Biology Program, Northeast Fisheries
Science Center, 166 Water Street, Woods Hole,
Massachusetts 02543, USA

Health supercourse to end Arab isolation

Sir — Your news feature "Blooms in the Desert" (*Nature* **416**, 120–122; 2002) focused on chinks of light against a gloomy backdrop of poor or nonexistent science funding. We, writing on behalf of the Islamic Global Health Network (islamicprevention.homestead.com), believe the critical element is the need to build Arabic scientific human capital, by building communities of Arab scientists.

Funding is important of course, but improving science training is more so, to unlock creativity and innovation. We in the field of epidemiology are in the process of making this real through the establishment of a global information-sharing Arab network, a model that can easily be followed for other scientific disciplines. The Internet is growing fast in Arab countries, so this can be used for collaborating and training, conserving scarce resources.

We are in the process of developing an Arabic 'Supercourse' of free-access PowerPoint lectures for instructors (www.pitt.edu/~super1/ighn.htm) as part of our global supercourse project, now consisting of more than 700 quality-controlled lectures from 118 countries (see *Nature Med.* **6**, 358, 2000). We are translating the Supercourse lectures into Arabic and providing special, continuously updated lectures precisely describing health issues in various Arab countries. So far, 130 Arab scientists are collaborating to translate, share lectures and build an Arabic scientific Supercourse.

Through Supercourse, we are trying to raise the quality of research and teaching in Arab institutions, ending the isolation of Arab scientists through greater Internet interaction with non-Arab countries. We want to bring brilliant young people from Arab countries into science, both locally and globally.

**Abdullatif Husseini, Rania Saad,
Ronald E. LaPorte**

Disease Monitoring and Telecommunications,
WHO Collaborating Center, and Graduate School
of Public Health, University of Pittsburgh,
Pittsburgh, Pennsylvania 15261, USA

Is a bell tolling for Bell Labs?

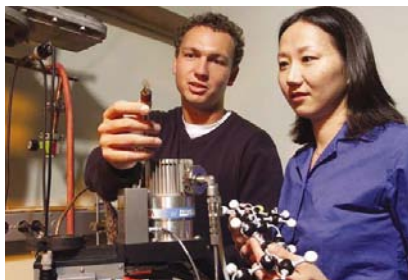
It would be wise of Bell Labs to help others reproduce their scientists' results.

Paul Grant

Dark clouds quickly began to gather over the exceptional finding of superconductivity at 117 K reported last year by Hendrik Schön and collaborators at Bell Laboratories in Murray Hill, New Jersey. Shortly after publication of the paper, I was asked by a reporter aware of my reputation as a sceptical observer of reports of 'unidentified superconducting objects' whether I felt uncomfortable that no one had reproduced any of the Bell Labs's field-effect transistor (FET) superconductivity results. My answer was: "Normally, I would be. But this is Bell Labs, and although these guys were my scientific adversaries for many years, I have the highest respect for their competence, credibility and, indeed, collegiality, and will accept their claims until proven otherwise."

Yet it has turned out that several attempts to reproduce not only these results, but also others on more general organic FET configurations, have been unsuccessful, culminating in allegations of duplication of data in several papers. Lucent/Bell Labs have set up a commission, headed by the internationally respected Malcolm 'Mac' Beasley, to examine the claims (see *Nature* **417**, 367–368; 2002 for an account of these events and for references). I hope this present imbroglio ends well for Bell Labs and its staff. For if not, we may indeed be hearing the final funeral tones from the halls of an already distressed US icon of science. To paraphrase John Donne, this bell will toll not only in Murray Hill, but throughout what remains of basic science research in industry.

Fabricating electrical contacts to conducting or semiconducting organic compounds is tricky. In the 1970s, in probably the first attempt to make organically based FETs, we at IBM's San Jose laboratory tried to form insulated-gate and Schottky barrier devices using lightly doped polyacetylene (CH_x), to make an internal thin-film transistor flat-panel display. The insulated-gate FET configurations were polyacetylene films deposited on oxidized degenerate silicon as a gate, and junction FETs formed by indium or gold contacts to semiconducting (CH_x). Both devices had weak transistor action, but not power gain as the 'barrier layers' were too conducting; yields were low and the lifetime of the barriers only a few days. After several attempts, we abandoned the effort—prematurely as it turned out, because a working organic FET was fabricated a few years later by a team at Cambridge (J. H. Burroughs, C. A. Jones & R. H. Friend, *Nature* **355**, 137–141; 1988). Several such devices now exist, including, presumably, those of the Bell Labs group.



Difficult times for Hendrik Schön and colleagues.

Perhaps even more germane to the difficulties of reproducing the results of Schön *et al.* were IBM's attempts to develop a thermal-transfer printing technology in the 1980s based on conducting polymer ribbons coated on one side with aluminium. Current was passed from a tungsten point contact on the bare side of the ribbon through the conducting polymer to the aluminium film, with the heat dissipated in the ribbon melting plasticized ink on the other side of the aluminium onto the paper. Print resolution was much better than we expected. It turned out that the interface between the aluminium and the conducting polymer was behaving as a Schottky-like barrier, which focused the current right underneath the point-contact print head. The interface between the aluminium and the conducting polymer was actually an unstable aluminium hydroxide layer displaying a highly non-linear current–voltage characteristic with a very high local electric field that sometimes quickly broke down and depended on ramping of the applied current. The hydroxide layer arose from minute amounts of water primarily responsible for the 10^{-6} torr background found in most vacuum systems used to deposit aluminium.

On 28 May this year, Schön circulated a preprint "Sputtering of alumina films for field-effect doping", apparently intended to guide people trying to reproduce his experimental conditions. Although his barriers are formed by sputtering Al_2O_3 and not by



Tried and tested: silicon-chip manufacture.

thermal evaporation of aluminium, the background pressure of the system is in the same 10^{-6} range where water vapour is the major component. The presence of aluminium hydroxide in plasma-deposited films of alumina has been documented (J. M. Schneider *et al. Appl. Phys. Lett.* **74**, 200–202; 1999) and is a source of some of the concern over the variability Schön reports in sample yield and lifetime. But it might also be the source of the magic that makes the devices work. There is a vague implication in Schön's preprint that only one sputtering unit has been used or has successfully produced functioning samples. If true, it would be a good idea to instrument that system to record all reasonably pertinent parameters in trying to build an exact copy.

The republic of science

Michael Polyani famously held that the practice of science is governed by a republic, a loose confederation of individuals bonded by the common goal of consensus on what determines when a particular 'truth' has been uncovered. In the republic of science, such truth is open to examination by everyone competent in whatever techniques are relevant to its determination. In more prosaic terms, if you're playing the sport of experimental physics, someone else had better reproduce your results or you're out of the ball game. If you've got an exciting result that may send you to Stockholm, the next thing to do, after you've established publication and patent priority, is to get your worst competitor to reproduce it, helping where necessary. Your management shouldn't have to set up a special commission to pore over your notebooks and files.

In the present instance, the issue is clearly one of samples. Traditional apparatus — batteries, potentiometers, lock-ins and x-y recorders — are all that is needed to measure electrical properties. Even if 'good' samples appear infrequently and are ephemeral when found, they can be taken to the lab next door and tested there. Even better would be transferring everything to another appropriate laboratory for replication. Setting up the Beasley commission is OK. But I believe an even more effective and decisive move would be for the leadership at Bell Labs to issue a 'grand challenge' to other suitable institutions to duplicate independently the extraordinary reports from their staff over the past two years, and to assist them to do so, thus showing that they are indeed good citizens of the republic of science. ■

Paul Grant is a science fellow at the Electric Power Research Institute, PO Box 10412, 3412 Hillview Avenue, Palo Alto, California 94303, USA.

Not waving but speaking

How important were gestures in the evolution of language?

From Hand to Mouth: The Origins of Language

by Michael C. Corballis

Princeton University Press: 2002. 384 pp.

\$27.95, £19.95

Michael Tomasello

The idea that communicative gestures played a prominent role in language evolution goes back several centuries. But the importance of gestures has not been given systematic treatment in the recent resurgence of interest in language origins, so this book by Michael Corballis is a welcome addition. But I must immediately add that, as is often the case in books that cover a wide range of material from several disciplines, the scholarship of the book is a bit lax in places — specifically, in those areas that I know something about — in ways that affect the theory in significant ways.

According to the theory, the common ancestor of chimpanzees and humans communicated in ways very similar to those of modern-day chimpanzees, using both vocalizations and gestures. Then, as humans set off down their own evolutionary pathway, they became bipedal, freeing their hands for developing gestures in richer ways. With the rise of the genus *Homo* some 2 million years ago, recursive thinking emerged, in which individuals think about thinking, based on adaptations to complex social problems. This led to a theory of mind — being able to think about what others are thinking. Recursion underlies complex syntax, so at this point humans had a complex language not unlike modern-day sign languages. Then, in the final step, vocalization took over from gesture, for a variety of reasons. This had several fortunate consequences, including a second freeing of the hands for other activities. Although this took place during the rise of the genus *Homo*, it was not fully complete until the arrival of modern humans — indeed, not until the final wave of migrations out of Africa some 50,000 years ago.

Obviously key to the theory is getting the starting point right. And this is one place where the book's scholarship leaves something to be desired. In characterizing the communication of our nearest primate relative, the chimpanzee, Corballis is correct to say that, in general, their gestures are learned and used more flexibly than their vocalizations, and to say that this is evidence in favour of a gestural origin for at least some aspects of language. But in developing this argument, he makes a number of errors.

Corballis is right to state that chimpanzee gestures do not refer to objects, but I



Talking point: unlike chimpanzees, human infants gesture towards objects simply to share attention.

disagree with his assertion that they “refer to actions” and are “iconic”. All the research that he cites demonstrates that chimpanzee gestures are not used to refer to outside entities at all, but simply to regulate social interactions between two individuals. The chimpanzee gestures listed in Table 3.1 of the book derive from social interactions and none is iconic. The claims for iconicity — which are disputed (though the disputes are not cited) — are made for gorillas only (and bonobos by a single researcher in a single context). Corballis also claims that chimpanzees conventionalize reaching for objects into a gestural indication for objects they want, but nothing like this has ever been observed in natural chimpanzee groups. In my view, chimpanzee gestures are learned primarily through ontogenic ritualization and not by emulation (except perhaps for some gestures that make a noise).

The outcome of these inaccuracies is that Corballis misses what is, in my opinion, the key difference between chimpanzee and human gestural communication. From a

very early age, human infants, unlike chimpanzees, point referentially to outside objects and events, and do so for no reason other than simply to share attention. These are both important features of human linguistic communication and lay the groundwork for the quintessential human use of language — idle conversation about external topics — and so are plausibly responsible for a major transition. But because he doesn't get the facts quite right, Corballis focuses on bipedal humans having their hands free more often as the crucial difference. This is only a quantitative difference, and I do not believe that it can account for the qualitative social-cognitive difference between human referential symbols (declarative pointing) and chimpanzee interaction-regulating imperative signals. For a discussion, see my book *The Cultural Origins of Human Cognition* (Harvard University Press, 1999).

There are a few other inaccuracies that also affect the argument. As part of his argument for the ‘naturalness’ of the manual modality, Corballis claims that children

learn sign languages more easily than spoken languages. But there is no citation here because there is considerable evidence (for example, the work of John Bonvillian and Virginia Volterra) that deaf children learn their first signs at the same age as hearing children learn their first words. Also in the domain of language acquisition, Corballis gives credit for the well-known book *Rethinking Innateness* (MIT Press, 1996) to Jeffrey Elman, Elizabeth Bates and Elissa Newport. The first two authors are correct, but the third is an outspoken opponent of the ideas in this book and will be horrified to see her name there (as will Mark Johnson, Annette Karmiloff-Smith, Domenico Parisi and Kim Plunkett, authors whose names are omitted).

Corballis also claims that Kanzi, a bonobo who understands some English words and phrases, could have done well in the comprehension experiments in which he participated by simply performing the most likely actions involving the objects involved. But the researchers conducting these studies dealt with this explicitly by having some experimental conditions in which semantic plausibility was controlled, and this is prominently underscored throughout the monograph.

Finally, Corballis cites as crucial support for his theory the existence of mirror neurons, which have been documented mainly for manual reaching, thus providing a neural substrate for learning gestures by imitation. But it is likely that if researchers were to look they would find the same thing for vocalizations — neurons that fire when an animal either produces or hears the same sound (call them echo neurons) — so the fact that neuroscientists looked first at reaching behaviour is not crucial. Unless we know that echo neurons do not exist, then mirror neurons do not provide crucial support for a gestural theory.

It is inevitable that any scholar who takes on a wide range of disciplines and tries to synthesize them into a coherent whole

cannot be an expert in all of them and is bound to be superficial in some places. In this case, there are several inaccuracies with the empirical data that in some cases lead to implausibilities in the theory. These could easily be rectified and would in some cases actually help Corballis' argument. But despite all this, he has still written an interesting and thought-provoking book that provides a welcome fresh perspective on the evolution of human linguistic communication. ■

Michael Tomasello is at the Max Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany.

Life out of nowhere?

Things Come to Life: Spontaneous Generation Revisited

by Henry Harris

Oxford University Press: 2002. 168 pp. £20, \$34.95

Paolo Mazzarello

One of the longest-lasting scientific problems in history is the question of spontaneous generation. The idea that animals and plants could be formed at any time from inanimate matter by a process of natural, spontaneous transmutation is rooted in the mists of time. But only in 1668 was it first tested scientifically, by Francesco Redi, who found no place for the spontaneous generation of macroscopic beings. The theory was resurrected some years later, after the discovery of microorganisms, which were considered to be a bridge between inorganic substances and life. In the following two centuries, some of the best natural scientists devised experiments that were intended to prove or refute the spontaneous development of life. But decisive experimental evidence against its occurrence was not produced until the



A need for seed? Louis Pasteur provided evidence that life could not develop spontaneously.

second half of the nineteenth century, by Louis Pasteur and John Tyndall.

The story of the crucial experiments and the alternate success and failure of this phoenix-like idea, resurrected and buried many times, is now clearly narrated by the cell biologist Henry Harris, author of *The Birth of the Cell* (Yale University Press, 1999), a history of cell theory. *Things Come to Life* is, in many respects, different from most other books on spontaneous generation. The origin of life is one of the fundamental questions that every culture has tried to answer in one way or another. For this reason, most studies of spontaneous generation deal mainly with the ideological controversy and its metaphysical meaning, rather than with the scientific details of the debate. Philosophy and theology are not entirely absent in this book (a brief chapter is entitled "Materialism, For and Against"), but its main aim is to describe the principal experiments and to judge what we can consider, in hindsight, to be right and wrong. So this is more a book of facts than of opinions about the facts, and more of experimental evidence than of the philosophical consequences.

Harris has avoided flooding the narrative with too many technical details and has added schematic reconstructions of some of the main apparatus used by the protagonists of this long story. For these reasons this small but informative book is suitable for a wide audience, including biologists (especially microbiologists), philosophers and historians of science. Because this is also the history of the rising of a modern biology postulate, according to which life comes from life and nothing is generated *ex nihilo*, it could be of formative value for students of medical and biological topics as well.

The story of the spontaneous-generation debate is a confirmation of the heuristic

New Journals

This year, *Nature's* annual new journals review supplement will appear in the 7 November issue. Publishers and learned societies are invited to submit journals for review, as well as details of any eligible electronic journals, taking note of the following criteria:

- Journals must have first appeared during or after June 2000 and published at least four separate issues by the end of June 2002.
- Journals covering any aspect of science are eligible, although those dealing with clinical medicine and pure mathematics are excluded, as are newsletters and publications of abstracts.
- Frequency of publication must be at least three times a year.
- The main language must be English.
- Deadline for submissions is 15 July 2002.

For each eligible title, please send at least four different issues (the first, the most recent and any two others), together with full details of subscription rates, to Mary Purton, *Nature*, The Macmillan Building, 4 Crinan Street, London N1 9XW, UK. Tel: +44 (0)20 7843 4567. Fax: +44 (0)20 7843 4596/4763. E-mail: m.purton@nature.com

importance that theories can have for the history of science, even when they are ultimately shown to be wrong. Wrong ideas do not always hinder the progress of science — they can even have positive consequences. Spurred on by the impossibility of solving the controversy with a single crucial experiment, scientists were engaged in a gigantic and often acrimonious hunt for biological findings in the hope of shifting the balance in one direction or another.

In the course of this battle, as Harris shows, many standard microbiological instruments were conceived, which influenced, and still influence, everyday laboratory life and the food industry (such as swan-necked flasks, cotton-wool air filters and autoclaves). So the wrong theory of generation without seeds was itself the seed of many ideas that had an enormous positive influence on the development of the microbiology and pathology of infectious diseases. ■

Paolo Mazzarelli is in the Dipartimento di Medicina Sperimentale, Sezione di Patologia Generale "C. Golgi", University of Pavia, Piazza Botta 10, I-27100 Pavia, Italy.

Gases get cool

Lévy Statistics and Laser Cooling: How Rare Events Bring Atoms To Rest

by François Bardou, Jean-Philippe Bouchaud, Alain Aspect & Claude Cohen-Tannoudji
Cambridge University Press: 2002. 214 pp.
£60, \$90 (hbk); £19.95, \$30 (pbk)

Bose-Einstein Condensation in Dilute Gases

by C. J. Pethick & H. Smith
Cambridge University Press: 2001. 414 pp.
£70, \$110 (hbk); £27.95, \$40 (pbk)

Keith Burnett

Laser cooling uses light fields to remove kinetic energy from atoms in the gas phase through the exchange of momentum between the light fields. The fact that this leads to cooling is remarkable in its own right. What is even more surprising, as was shown in pioneering work in which the authors of *Lévy Statistics and Laser Cooling* played a pivotal role, is that atoms can be made to randomly pick up then throw away the momentum gained from the field until they find themselves 'trapped' in a region of low momentum.

This process can be used to cool atoms to low temperatures (sub-microkelvin, a fraction of a degree above absolute zero) where an atom's momentum is less than that of a single photon, a technique known as sub-recoil laser cooling. This method of laser-cooling atoms, known as velocity-selective coherent population trapping (VSCPT),

means that atoms can be cooled to close to the conditions for producing Bose-Einstein condensates. These condensates are unique sources of coherent matter waves and are opening up new avenues of research in physics, as discussed in the second book. But more of that later.

Laser cooling has led to a broad range of new theories and experiments, as well as the 1997 Nobel Prize in Physics for Claude Cohen-Tannoudji, one of the authors of this book. The crucial issue, from the point of view of the book, is the extraordinary utility of the method of Lévy statistics for qualitative and quantitative understanding of laser cooling based on VSCPT. The authors have provided an excellent and readable account that will be of considerable use not only to people interested in laser cooling, but also to those wishing to see this important set of techniques make an impact in studies of ultracold matter.

The authors first give a brief explanation of the methods of laser cooling, with the details of the link between the physics and the models discussed in the book being partly relegated to an appendix to avoid breaking the flow of the text. They then show how the process calls out for a statistical method that can handle broad distributions. Lévy statistics, which is also used in many other fields including biology, Earth sciences and finance, fits the bill. This statistical method is described in just the right amount of detail

for the reader to appreciate its use in laser-cooling theory. The application of the general concepts to the problem at hand is then given in clear and convincing terms, supported by a solid, physically based discussion of what is going on. The book then gives more specifics of the results, which will be used eagerly by other workers in the field. This includes, most importantly, methods for optimizing the cooling produced for various systems that can be used in the laboratory.

The book is a significant addition to the literature in both laser cooling and statistical physics. It is rare to have such a lucid and convincing account of a technique that will be new to most scientists. It will be greatly welcomed both by workers in the field of ultracold atom physics and by those who want to see an important theoretical apparatus used in practice.

One of the most exciting applications of laser cooling is the production of atoms that are sufficiently cold and dense to be evaporatively cooled to produce Bose-Einstein condensation. The laser-cooled atoms are trapped using magnetic fields and cooled by evaporation of the more energetic atoms. *Bose-Einstein Condensation in Dilute Gases* is an excellent and much-needed text of the theory of these condensates.

In a condensate, all the atoms in a gas share exactly the same de Broglie wave. In this sense they constitute giant quantum waves, in which the wave nature of matter



Look back in amber

Republished in English for the first time, the *Atlas of Plants and Animals in Baltic Amber* by Wolfgang Weitschat and Wilfried Wichard (Friedrich Pfeil, 75 euros, \$98) is a rich catalogue of the flora and fauna that have

been preserved in Baltic amber. It guides us through the history and geography of petrified resin and discusses the geology of amber. This spider is one hundreds of life-forms captured in amber that illustrate the book.

is writ large as the wavefunctions approach macroscopic, millimetre size. One can also think of condensates as being laser-like sources of matter waves, all the atoms having the same wave nature and state of motion. They are coherent matter wave sources or 'atom lasers' in the same sense that ordinary lasers are coherent sources of light wave photons. Because of these unique properties as macroscopic quantum systems, they are finding applications in such diverse areas of science as superfluidity, quantum computing and precision measurements based on atom interferometry.

Bose–Einstein condensates were first formed in the laboratory in 1995 by using evaporative cooling of gases to temperatures in the nanokelvin range. This work opened the door to a wide range of new physics. The pace of advances has been quite breathtaking, with studies of the physical properties of the condensates and, most recently, the prospect of their use in quantum-computational schemes. The 2001 Nobel Prize was awarded to Carl Wieman, Eric Cornell and Wolfgang Ketterle for their spectacular achievements in this field. Although progress continues at a cracking pace, there is now a set of basic notions that it is sensible to teach postgraduates, including the way that condensates are made and their physical properties as macroscopic quantum systems. This book is an excellent source of information on this topic, and is accessible to a wide range of physicists and chemists.

Previous accounts of Bose–Einstein condensation are scattered among various papers and review articles. Older texts that deal with the physics of Bose gases or superfluids cover some of the issues but deal with matters that are relevant to the new experimental scene in a cursory manner. So this is a most welcome text for those of us wishing to expound the new physics to a younger generation of physicists. The range of topics and level of detail is well matched to the needs of someone with little background in atomic or optical physics. Topics covered include the microscopic theory of trapped condensates and their superfluid properties, excitations and vortices. The authors have also included problems at the end of each chapter, enhancing the book's value as a teaching aid.

It is inevitable in such a rapidly moving field that the book does not contain information on the most recent advances concerning Bose–Einstein condensation in lattices, quantum phase transitions and quantum information processing (for reviews see *Nature* **416**, 205–246; 2002). These will probably be in the next edition of what is likely to be a best seller in its category. This well-produced book is a 'must buy' for anyone wanting to get started in this field. ■

Keith Burnett is in the Department of Physics, Clarendon Laboratory, University of Oxford, Oxford OX1 3PU, UK.

Science in culture

Projecting an image

Did the Old Masters paint using optical projection techniques?

Michael John Gorman

Painters from Jan van Eyck to Jean-Auguste-Dominique Ingres may have achieved a remarkable imitation of nature not through sheer painterly talent, but by using optical devices, according to David Hockney's controversial recent book *Secret Knowledge: Rediscovering the Lost Techniques of the Old Masters* (Penguin Putnam, 2001). Specifically, Hockney, assisted by optical scientist Charles Falco of the University of Arizona, argues that from around 1430 many painters used a concave mirror to project brightly lit subjects onto a canvas, allowing them to render figures with unprecedented naturalism. Later, at around the end of the sixteenth century, according to Hockney and Falco, painters including Caravaggio began to use refractive lenses instead of concave mirrors to project their images for tracing (see *Nature* **412**, 860; 2001).

Since the publication of Hockney's provocative thesis, several objections have been raised. For example, it has been claimed that Hockney's projection technique would have required a mirror of very long focal length to create an image suitable for tracing, and that this would exceeded the technical capabilities of Renaissance mirror-makers.

New evidence concerning the history of optical projection further challenges Hockney's hypothesis. In particular, a close reading of the historical documents suggests that the specific device that Hockney and Falco claim was used widely by artists from the 1430s was in fact invented by the Neapolitan magician Giambattista della Porta in 1558, and then rendered obsolete by della Porta himself in 1589.

Della Porta was as famous for his investigations of arcane natural processes and mechanical contrivances as for being a playwright and impresario. One of the many instruments that intrigued him was the camera obscura — the generic name given to the projection of inverted images through a small hole into a darkened chamber. In 1558, he gave the earliest description of a new type of camera obscura in the first edition of his widely read book *Natural Magic*. The new technique involved using a concave mirror to project an inverted image onto a piece of paper. This is the first documented account of the device that Hockney and Falco claim was used by artists from the 1430s.

Incidentally, the first account of incorporating a convex lens into the camera obscura dates from just eight years earlier, in the encyclopedic work of the astrologer and mathematician Girolamo Cardano called *On Subtlety*, which also describes in detail the workshop techniques of contemporary painters.



Caravaggio may have used della Porta's camera obscura when painting *The Calling of St Matthew*.

In the expanded second edition of *Natural Magic*, published in 1589, della Porta added a dramatic revision to his concave-mirror camera obscura. As a result of his extensive investigations of optical instrumentation on the Venetian glass-making island of Murano in 1580, he combined a convex lens with the concave-mirror projection system. The remarkable result was a device that projected large, upright images. The inverted image formed by the lens, falling a short distance in front of the focal point of the concave mirror, served as an object for the concave mirror, which turned it upright and magnified it. In this way, even a concave mirror of short focal length, within the manufacturing capabilities of the late sixteenth century, could be used to project life-sized images into a darkened room, given an appropriate lens. Della Porta used this device, which doubled as a primitive reflecting telescope, not to paint but to project extravagant theatrical performances for aristocratic audiences seated in his darkened chambers.

Where does this leave the earlier paintings discussed in Hockney's book? At best, the device that Hockney claims was used by artists from the 1430s had perhaps a 35-year working life as an artist's instrument over 100 years later, assuming that it was not just an amusing toy for those unable to draw, as della Porta himself suggests. If Caravaggio used a camera obscura, he would have had every opportunity to avail himself of the latest technology — della Porta's combined convex lens and concave mirror, an instrument that goes unmentioned in Hockney's book. If fifteenth-century painters ever used the camera obscura, however, then they used a simple hole in the wall: no mirror, no lens.

Michael John Gorman is in the Program in Science, Technology and Society, Stanford University, Stanford, California 94305, USA.

◆ <http://www.stanford.edu/group/shi/eyes/hockney>

ARALDO DE LUCA/CORBIS

Strength in numbers

Austin L. Hughes

Every species has a history. Although no chronicler was there to record them or bard to immortalize them, fruitflies, flycatchers and kangaroos have had their epic migrations, their great plagues, their eras of prosperity and population growth, and their periods of decline.

For our own species, we give the name 'prehistory' to all of our history that occurred before written records. In this sense, most of the history of other species is also prehistoric, and until quite recently we had little means of knowing about it. Only a few of our fellow life-forms — such as domestic animals and our major pests — have found their way into our written records or left traces at our archaeological sites. Most non-human species have lived their histories away from us and have left no artefacts for us to find out about them. Our prospects of reconstructing much of our own prehistory used to be little better. Archaeological evidence is often ambiguous — for example, when technologies or styles of art change, was this because new techniques had been learned or because another culture had replaced the previous one? Even the written record poses as many historical mysteries as it resolves, as we cannot distinguish

between propaganda and objective accounts.

Molecular-biology data offer the promise of at last unlocking the prehistories of our own and other species. In the case of ourselves and a few other organisms, we now have an invaluable resource — a complete genome sequence. Using this sequence as a guide, we can re-sequence from different populations a substantial number of loci sampled from throughout the genome.

Should we believe it when we are told that genetic markers can be used to reconstruct all sorts of ancient events, from the origin of modern *Homo sapiens*, to the spread of agriculture, to the peopling of oceanic islands? It is true that genetic markers — unlike ancient chroniclers — do not lie. But their interpretation raises many a thorny problem and can be as perilous as attempts to decipher ancient inscriptions in an unknown tongue.

Consider the most straightforward type of question that we might want to answer about a species' history. Suppose we want to date the separation of two populations of the same species, divided by a geographical barrier that prevents or substantially reduces gene flow between them. The separation might correspond to a major migration — for example, the first migration of modern humans out of Africa. Naively, we might assume that answering this question would be very simple. We could sequence alternative forms of a gene (alleles) from an individual in each of the two populations and count the base-pair differences between the two sequences. Given an estimate of the mutation rate and the assumption that this rate is roughly constant over time, it is straightforward to estimate the age of the common ancestor of the two sequences.

But this date does not necessarily correspond to the most recent common ancestor of the two populations, because the ancestral population itself would have contained genetic differences. Just by chance, an allele that existed in the ancestral population might have ended up being fixed in one of the descendant populations, while another allele was fixed in the other descendant population. The common ancestor of these two alleles will be significantly more ancient than the time at which the two populations separated, especially when balancing selection acts to maintain such a polymorphism.

In vertebrates, the best-studied case of an ancient, balanced polymorphism is the highly polymorphic loci of the major histocompatibility complex (MHC) of the immune system, in humans generally termed HLA (for human leukocyte antigen). Polymorphic alleles at HLA loci — or, more accurately, allelic lineages — can be very ancient, even pre-dating the most recent common ancestor of humans

Genetic markers

Genetic markers — unlike ancient chroniclers — do not lie. But their interpretation can be as perilous as attempts to decipher ancient inscriptions in an unknown tongue.

and chimpanzees. It now seems that balancing selection at other loci in the human genome may be more common than previously thought — there is probably a balanced polymorphism at the dopamine-receptor D4 locus that could relate to behavioural differences between people with different alleles for this gene. The melanocortin 1 receptor locus, which is involved in pigmentation of skin and hair, is another surprisingly polymorphic site at which balancing selection may operate.

The obvious solution is to sample as many loci as possible. (It is worth noting in this context that the entire mitochondrial genome — the workhorse of genetic marker studies to date — represents only a single locus, as it is inherited as a unit without recombination.) If gene flow between two populations has been completely eliminated, the set of loci that shows the most recent ancestor for the two populations must correspond to loci that were monomorphic at the time of their separation. However, in many cases, gene flow between isolated populations is reduced but not eliminated. Population migrations may follow episodic patterns, as in the multiple migrations out of Africa over hundreds of thousands of years that have recently been proposed for our own species. Given such a complex history, there is a real danger that polymorphism pre-dating population subdivision will be taken as evidence of a still more ancient migration.

Data on worldwide patterns of sequence polymorphism at hundreds or even thousands of loci provide the statistical power to discriminate between chance patterns observed at one or a few loci, and can hence reveal true population histories. In the world of genetic markers, there is definitely strength in numbers. ■

Austin L. Hughes is professor of biological sciences and director of the Biotechnology Institute, University of South Carolina, Columbia, South Carolina 29208, USA.

FURTHER READING

- Ding, Y. C. *et al. Proc. Natl Acad. Sci. USA* **99**, 309–314 (2002).
 Harpending, H. C. *et al. Proc. Natl Acad. Sci. USA* **95**, 1961–1967 (1998).
 Takahata, N. *et al. Mol. Biol. Evol.* **18**, 172–183 (2001).
 Nei, M. *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York, 1987).
 Templeton, A. R. *Nature* **416**, 45–51 (2002).



Tell-tale signs: like this Bronze Age tablet, genetic polymorphisms can offer clues to human history.

Teamed up for transcription

Christian von Mering and Peer Bork

In bacterial genomes, functionally related genes are often clustered and controlled as a unit. Such 'operons' are not normally found in animals — so why are they so abundant in one class of worm?

The genomes of animals, plants and fungi seem to be relatively disorganized. Genes appear to be randomly distributed, with only a few exceptions: repeats of similar sequences caused by gene duplications, for example, and a limited number of ancient gene clusters containing functionally related genes (such as the Hox genes that are involved in control of animal development¹). Apart from these, the average gene is generally assumed to be independent of its neighbours, and genomes are constantly rearranged and shuffled. However, in one group of animals — the nematodes (small, unsegmented worms) — neighbouring genes are occasionally assembled into regulatory units called operons². On page 851 of this issue, Blumenthal *et al.*³ now report the first whole-genome characterization of such operons in a multicellular organism, and raise intriguing questions as to how (and why) they have evolved.

Operons were first described⁴ in prokaryotes — small, single-celled organisms without a nucleus. In these organisms, genes needed for one particular process (say, synthesis of an amino acid) are often clustered in close proximity on the genome, with the same orientation on the DNA. These operons facilitate the coordinated regulation of genes, as the clustered genes are activated (that is, transcribed into messenger RNA) all in one go. Operon transcripts always code for more than one protein, and prokaryotes handle this by starting translation of the messenger RNA into amino acids separately at the beginning of each protein-coding section (Fig. 1a). In contrast, these 'polycistronic' transcripts are a problem for eukaryotes (organisms with nucleated cells, including all multicellular organisms). They generally cannot process such transcripts, and so it has long been assumed that most eukaryotes do not have operons.

Blumenthal *et al.*, however, had shown previously² that the nematode *Caenorhabditis elegans* occasionally does make polycistronic transcripts and can process them through a mechanism called *trans*-splicing. Briefly, the transcripts are split up into individual pieces, and short 'leader' sequences are simultaneously added to each piece (Fig. 1b). In *C. elegans*, *trans*-splicing is not limited to polycistronic transcripts — in fact, the majority of transcripts receive leader sequences. But the polycistronic transcripts are unique in

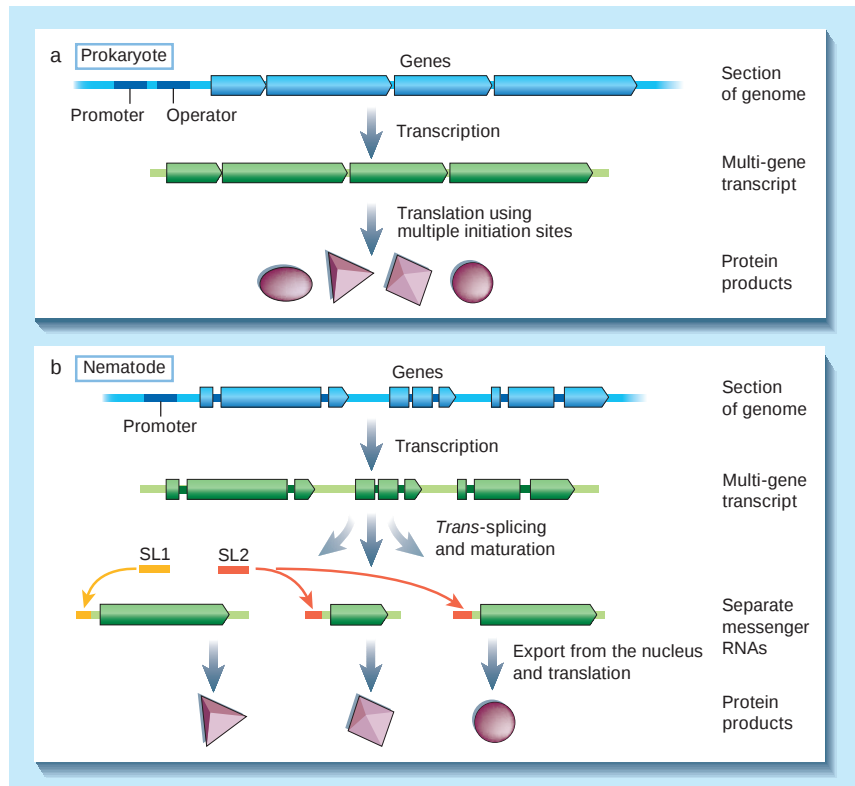


Figure 1 Operons — organizing genes for joint activation. a, A classical operon, found in prokaryotes (such as bacteria). Several genes, often functionally related, form a tight cluster on the genome. Operons are under the control of regulatory elements (promoter, operator) and factors that bind to these elements. Transcription of the cluster results in a single molecule, a multi-gene transcript of messenger RNA, which codes for several proteins and is directly translated into distinct protein products. b, An operon in the nematode *Caenorhabditis elegans*, typical of the several hundred discovered by Blumenthal *et al.*³ in the first genome-wide survey for operons in a eukaryotic organism. The initial multi-gene transcript is split up into separate messenger RNAs by *trans*-splicing, concomitant with the insertion of splice leader sequences (SL1, SL2). Although these operons fit the classical genetic definition⁴, the processing of their transcripts into proteins differs fundamentally from that of prokaryotic operons.

that processing involves a particular splice leader, called SL2, for all pieces but the first (which often has the usual leader, SL1).

To measure the extent of polycistronic transcription (and so of operons) in *C. elegans*, Blumenthal *et al.*³ initiated a systematic, genome-wide study. Reasoning that it should be possible to discover all operons by searching for SL2 transcripts and checking their arrangements on the genome, they took two independent approaches. First, they analysed *C. elegans* transcripts that had been sequenced previously (random transcripts are routinely sequenced in many model organisms). Among these, they found more

than 300 that contained SL2. For their second approach, they prepared a collection of actual transcripts from a population of worms, and enriched the sample for transcripts with an SL2 leader (this involved adding synthetic SL2-like molecules to the mixture to find the right transcripts through base-pairing). The enriched population was then compared to a normal, non-enriched fraction by measuring transcript abundances. This technically demanding approach yielded another 1,200 SL2-transcripts. Both sets of data converged well and agreed with gene clustering as observed in the genome. Together with the unlabelled

first gene of each operon, the combined list contains 2,291 genes in 881 operons. This is a surprisingly large total — the authors estimate that 13–15% of all genes in *C. elegans* (including some they might have missed) are expressed as part of an operon.

What can be learned from this first genome-wide study of eukaryotic operons? First, it nicely confirms a previous assumption²: in *C. elegans*, SL2-containing transcripts are indeed reliable operon markers (as expected, in many runs of consecutive genes on the genome, all genes but the first were found to generate SL2-transcripts). Second, the data may help researchers working with an uncharacterized gene — if it is in an operon with a gene that is better characterized, they can probably assume that both have the same transcriptional regulation. Third, a genome-wide view of operons in *C. elegans* is a step towards addressing long-standing questions such as why nematodes have operons and whether they are the only animals that do.

Blumenthal *et al.* cautiously suggest that some of the operons could serve the same purpose as their counterparts in prokaryotes: to group functionally related genes together. This clearly appears to be true for some genes³, and it would be exciting if it were the case for most of them — it would mean that the function of uncharacterized genes could be predicted from their operon context. We could not resist checking the functional annotations of these newly discovered operons (using data from the Gene Ontology Consortium⁵), but we found that only about 4% of the operons contained two or more genes annotated for the same biological process, compared to 36% for known and putative operons in the prokaryote *Escherichia coli*. This figure is low, but it could well be due to the fact that the fraction of genes with known function is smaller in *C. elegans* than in prokaryotes. Because of these limits in annotation, we cannot yet say whether most genes in operons are functionally related — but there are certainly cases where they are not.

To understand the functional significance of operons in *C. elegans*, one must consider their evolutionary history. Are they related to prokaryotic operons through common descent? *Trans*-splicing, at least, is not unique to nematodes. It was first described in single-celled eukaryotes (trypanosomes, in which entire operons have also been reported⁶), and more recently in several animals other than nematodes^{7,8}. This raises the possibility that the earliest multicellular animals possessed *trans*-splicing capability, and possibly had operons inherited from prokaryotes. However, in addition to mechanistic differences (Fig. 1), the content of nematode operons seems distinct from that of prokaryotes. Not only is the average number of genes per operon smaller (2.6 in *C. elegans* against 4.1 in *E. coli*), but we also found only a few cases where two or

more genes were in an operon together in both *C. elegans* and prokaryotes — hardly more than would be expected by chance (using the database Clusters of Orthologous Groups⁹). Furthermore, no operons have yet been found in well-studied groups such as insects, fishes and mammals, nor in plants or fungi.

From all of this it seems that operons in *C. elegans* were a lineage-specific invention, perhaps facilitated by an existing capability for *trans*-splicing. A possible evolutionary scenario is suggested by the finding that genes with similar transcription profiles may have a tendency to cluster in eukaryotic genomes^{10,11}. It is conceivable that such clusters were among the first to be joined into operons. This might have been promoted by a selective pressure for a small genome (*C. elegans* is known for its compact genome). Later, in the continuing processes of genome shuffling, segmental duplications and gene loss, more transcriptionally or functionally related genes could have come close together by chance, forming additional operons.

Whatever the precise course of events,

Blumenthal *et al.*³ have provided us with a key study. The operons they have identified will be the basis for much further investigation into genome evolution and gene regulation in animals.

Christian von Mering and Peer Bork are at the European Molecular Biology Laboratory, Meyerhofstrasse 1, 69012 Heidelberg, and the Max-Delbrück Centre for Molecular Medicine, Berlin, Germany.

e-mail: bork@embl-heidelberg.de

1. Ferrier, D. E. & Holland, P. W. *Nature Rev. Genet.* **2**, 33–38 (2001).
2. Spieth, J., Brooke, G., Kuersten, S., Lea, K. & Blumenthal, T. *Cell* **73**, 521–532 (1993).
3. Blumenthal, T. *et al.* *Nature* **417**, 851–854 (2002).
4. Jacob, F. & Monod, J. *J. Mol. Biol.* **3**, 318–356 (1961).
5. The Gene Ontology Consortium *Genome Res.* **11**, 1425–1433 (2001).
6. Gao, G., Nara, T., Nakajima-Shimada, J. & Aoki, T. *J. Mol. Biol.* **285**, 149–161 (1999).
7. Stover, N. A. & Steele, R. E. *Proc. Natl Acad. Sci. USA* **98**, 5693–5698 (2001).
8. Vandenbergh, A. E., Meedel, T. H. & Hastings, K. E. *Genes Dev.* **15**, 294–303 (2001).
9. Tatusov, R. L. *et al.* *Nucleic Acids Res.* **29**, 22–28 (2001).
10. Cohen, B. A., Mitra, R. D., Hughes, J. D. & Church, G. M. *Nature Genet.* **26**, 183–186 (2000).
11. Lercher, M. J., Urrutia, A. O. & Hurst, L. D. *Nature Genet.* **31**, 180–183 (2002).

Astronomy

Nebulous explanation

Mark Claussen

The planetary nebulae of gas and dust that are formed from red-giant stars are usually far from spherical in shape. Observations of the gas distribution in one red giant caught in the act of transition show why.

Planetary nebulae are glowing remnants of gas, left over from the dying stages of Sun-like stars. Images from the Hubble Space Telescope have provided astronomers with exquisitely detailed images of these beautiful objects. But a long-standing mystery is why most (but not all) planetary nebulae are not spherical, although the distributions of gas and dust in their immediate predecessors are. Now Imai *et al.*¹ (page 829 of this issue), using a ground-based interferometric technique, provide evidence that a well-collimated outflow, or jet, of gas is emanating from a red-giant star that is in transition to becoming a planetary nebula. Their observations also suggest that the jet may be precessing — wobbling around on its axis like a toy top does as it spins. Their findings should lead to a better understanding of how planetary nebulae are shaped.

The Sun is a fairly common type of star, and will remain substantially unchanged over the next several billion years. Nuclear fusion reactions (converting hydrogen into helium) in the core of the Sun provide energy that supports the outer gaseous layers as gravity tries to force them inwards. But after many billions of years, when all the hydrogen in the core of such a star has been converted into helium, the core contracts under gravity,

heating up and creating pressure that causes the outer layers of the star to expand and cool. The star becomes a red giant; when the Sun reaches that stage, its outer layers will expand as far as the Earth's orbit.

As the expansion progresses, the outer layers of the star become unstable and the star pulsates, changing its total energy output on a regular basis (the cycle time is usually around a year). The pulsation causes matter to be lost from the star as a very slow 'wind' (with a velocity of about 10 km s⁻¹), into a surrounding shell. These pulsational changes last several tens of thousands of years. Eventually, the core of the star collapses to form a white dwarf with a surface temperature of 30,000 K — hot enough to ionize the gas in the shell that surrounds the dying star. The ionized gas is what we see as a planetary nebula.

This model of the later stages of stellar evolution works just fine until the last part. The slow mass loss is, as far as we can tell, generally spherically symmetrical. So the distribution of any gas that is expelled in this manner should have spherical symmetry. And yet the ionized gas distribution for most planetary nebulae is far from spherical (Fig. 1); many have a flattened, even two-lobed, structure^{2–5}.

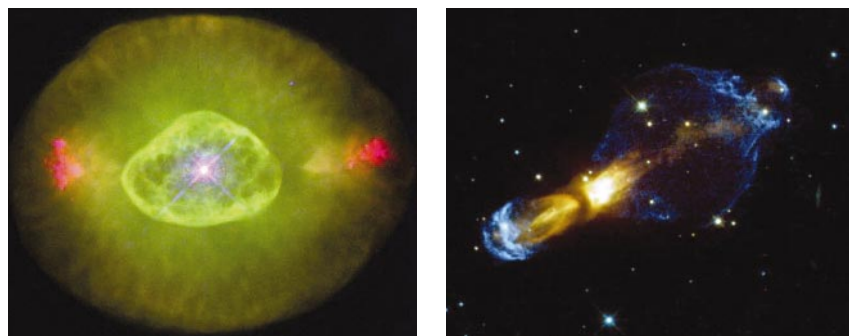


Figure 1 The dying of the light. These Hubble Space Telescope images show planetary nebulae NGC 6826 (left) and OH 231.8 (right) — the gaseous remnants surrounding dying stars. Gas expelled from the stars during their red-giant phase forms a spherical shell around the stellar core. But most planetary nebulae are not spherical. New observations by Imai *et al.*¹ of the planetary nebula W43A reveal a strong, precessing jet of gas escaping from its core. This may explain the loss of spherical symmetry; moreover, the motion of the jet suggests that W43A has been caught exactly at the point of its transition from red giant to planetary nebula.

During the red-giant stage, molecules form in the gas that has been expelled into the circumstellar shell. The presence of molecules such as hydroxyl, silicon monoxide and water has been established from their maser emission⁶ (the microwave analogue of laser emission). The study of maser emission at radio wavelengths is a useful tool for two reasons. First, the gas motion in the shell can be determined from the Doppler-shifted frequency of emission. And second, the masers are so bright that a ground-based technique, ‘very long baseline interferometry’ (VLBI), can be used to measure their position accurately. VLBI combines the measurements of several telescopes positioned over a wide area. In the case of water masers, their position can be established to better than 6×10^{-8} degrees using VLBI. This technique has become routine since the opening in 1992 of the National Radio Astronomy Observatory’s Very Long Baseline Array (VLBA), a dedicated VLBI instrument of ten radio antennae extending from Hawaii to the Virgin Islands.

As the central stellar core becomes hot enough to ionize the surrounding gas, the molecules in the gas are expected to be destroyed, so their maser action should cease. But Miranda *et al.*⁷ have found water and hydroxyl masers in a very young planetary nebula, and Likkell *et al.*⁸ discovered water masers in W43A, the nebula observed by Imai and colleagues¹.

W43A is an intriguing object, thought to be in transition from a red giant to a planetary nebula. Imai *et al.* have used the VLBA to determine the distribution of water masers in the nebula with a precision 200 times greater than that of optical observations by the Hubble Space Telescope. Surprisingly, they found that the gas traced by the maser emission was not spherical in distribution, but instead forms rather narrow, oppositely directed jets. The gas in the jets is moving at 150 km s^{-1} — although that is

not particularly high for gas jets, it does lead to an estimated age for the jet of only around 30 years. This strongly suggests that the star has indeed been caught in the act of transition to a planetary nebula.

Imai *et al.* also draw another significant conclusion: the masers do not lie in a straight line but on a curve, suggesting that the source of the outflow might be precessing. This is perhaps the most exciting part of their observations. There are several possible explanations for such a precessing jet — perhaps a

close binary companion⁹, or strong magnetic fields around the star¹⁰.

The molecular jet seen by Imai *et al.* suggests a mechanism for producing the non-spherical morphologies seen in most planetary nebulae, and further observations of young planetary or transition objects promise to finally solve the mystery. The next steps, necessarily a collaborative effort between theorists and observers, will be to investigate what causes the jet, to confirm its precession and to determine how the precession occurs. ■

Mark Claussen is at the National Radio Astronomy Observatory, PO Box 0, 1003 Lopezville Road, Socorro, New Mexico 87801, USA.

e-mail: mclaussen@nrao.edu

1. Imai, H., Obara, K., Diamond, P. J., Omodaka, T. & Sasao, T. *Nature* **417**, 829–831 (2002).
2. Balick, B. *Astrophys. J.* **94**, 671–678 (1987).
3. Kwok, S. in *Asymmetrical Planetary Nebulae II: From Origins to Microstructures* (eds Kaster, J. H., Soker, N. & Rappaport, S. A.) **199**, 9–16 (2000).
4. Terzian, Y. & Hajian, A. R. in *Asymmetrical Planetary Nebulae II: From Origins to Microstructures* (eds Kaster, J. H., Soker, N. & Rappaport, S. A.) **199**, 33–40 (2000).
5. Balick, B. in *Asymmetrical Planetary Nebulae II: From Origins to Microstructures* (eds Kaster, J. H., Soker, N. & Rappaport, S. A.) **199**, 41–48 (2000).
6. Reid, M. J. & Moran, J. M. *Annu. Rev. Astron. Astrophys.* **19**, 231–276 (1981).
7. Miranda, L. F., Gómez, Y., Anglada, G. & Torrelles, J. M. *Nature* **414**, 284–286 (2001).
8. Likkell, L., Morris, M. & Maddalena, R. J. *Astron. Astrophys.* **256**, 581–594 (1992).
9. Morris, M. *Publ. Astron. Soc. Pacif.* **99**, 1115–1122 (1987).
10. Garcia-Segura, G. *Astrophys. J.* **489**, L189–L192 (1997).

Physiology

Two ACEs and a heart

Kenneth E. Bernstein

Cardiovascular diseases are some of the biggest killers in the developed world. The discovery of a new enzyme that affects cardiac function might provide fresh insight into heart disease.

For years now, scientists have thought that they understand the formation, function and fate of the eight-amino-acid peptide angiotensin II. This peptide is the result of the progressive degradation of a precursor protein, angiotensinogen, by the enzymes renin and angiotensin-converting enzyme (ACE). Once formed, angiotensin II is a master regulator of body physiology, raising blood pressure by coordinated effects on the heart, blood vessels and kidneys. The importance of angiotensin II is also implied by the wide clinical use of ACE inhibitors — drugs that reduce the formation of angiotensin II — to treat high blood pressure, heart failure and even some of the destructive effects of diabetes. But, as Crackower and colleagues make clear on page 822 of this issue¹, the ‘renin–angiotensin’ system is even more complicated than was thought. It seems that the recently discovered enzyme ACE2 also has a say in how the heart functions.

The story begins in 2000, with the discovery^{2,3} of ACE2 as an enzyme similar to ACE — both are carboxypeptidases, meaning that they catalyse the shortening of peptides from their carboxy-terminal ends. But whereas ACE cleaves two amino acids at a time, ACE2 shortens peptides by only one amino acid. The upshot is that angiotensin peptides come in several slightly different sizes, yet have very different effects. Angiotensin I (with ten amino acids) is an intermediate peptide without significant biological effects (Fig. 1, overleaf). Angiotensin I is thought to be converted to angiotensin 1–9 (with nine amino acids) by ACE2, but to the eight-amino-acid angiotensin II by ACE. Whereas angiotensin II stimulates blood-vessel constriction and raises blood pressure, angiotensin 1–9 has no known effects. Angiotensin 1–9 cannot be converted to angiotensin II by ACE2, but can be converted to angiotensin 1–7 (a blood-vessel dilator) by ACE. So it has

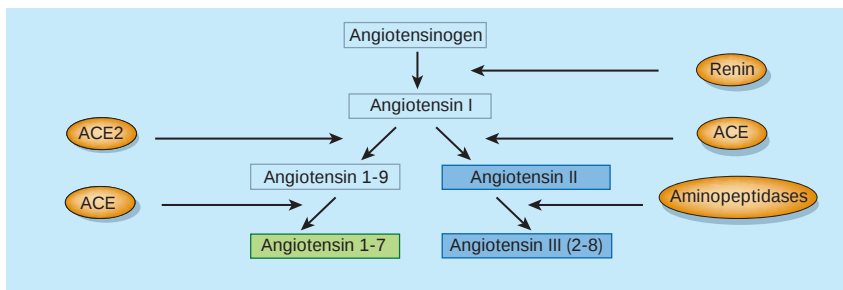


Figure 1 Angiotensin peptides. Blood-vessel constrictors are in blue; vessel dilators are in green. The enzyme renin catalyses the release of the ten-amino-acid peptide angiotensin I from the precursor protein angiotensinogen. Further catalysis by ACE produces the eight-amino-acid angiotensin II, which is a powerful vessel constrictor and results in higher blood pressure. ACE2 is thought to produce the nine-amino-acid angiotensin 1-9, which has no known effects but can be converted by ACE to angiotensin 1-7, a vessel dilator. So ACE2 has been proposed to reduce the formation of angiotensin II and to oppose the increase in blood pressure that is mediated by ACE and angiotensin II. Crackower *et al.*¹ find that the ACE2 gene is found on a region of the rat X chromosome that has been implicated in high blood pressure. Yet knocking out ACE2 in mice does not raise blood pressure, but instead leads to malfunctioning heart muscle.

been suggested that ACE2 prevents the formation of the blood-vessel constrictor angiotensin II.

But what do the effects of ACE2 on angiotensin peptides mean in terms of physiology? A first hint was the finding that this enzyme is expressed solely in the heart, kidneys and testes^{2,3}. But it is the paper by Crackower *et al.*¹ that provides the first really compelling evidence of ACE2's biological

role. First, the authors present gene-mapping studies designed to find the location of the ACE2 gene in the rat genome, specifically on the rat X chromosome. The reason was that studies of several rat 'models' of high blood pressure had suggested that the X chromosome contains a particular region (perhaps even a particular gene) that is important in the heritability and induction of raised blood pressure. Indeed, Crackower *et al.* found that

the ACE2 gene maps to the same region of the X chromosome that has been implicated in three such models. Moreover, ACE2 protein levels were reduced in these models.

So is ACE2 involved in controlling blood pressure? To find out, Crackower *et al.* created 'knockout' mice lacking all ACE2 protein. Mice lacking the ACE protein have very low blood pressure^{4,5}. Thus, one might imagine that knocking out ACE2 has the opposite effect — high blood pressure — as suggested by the rat models. In fact, the authors found that blood pressure is normal in the ACE2-knockout mice. But surprisingly, and unlike the ACE-deficient mice, the ACE2-knockout animals develop abnormal (pathological) heart function as they age. Specifically, the heart muscle develops a significant defect in both the speed and the overall percentage of contraction. Unusually, though, the hearts show only mild physical changes, with virtually no increase in cardiac scarring.

Why don't the hearts of the ACE2-knockout mice function properly? Crackower *et al.* found that the lack of ACE2 leads to an increase in angiotensin II levels, and suggest that this may, in part, incite the pathology (through heart-specific mechanisms that apparently would not involve an effect on blood pressure). It is wise to keep an open mind, however. The increased levels of angiotensin II are significant but relatively

Zoology

New whale from old bones

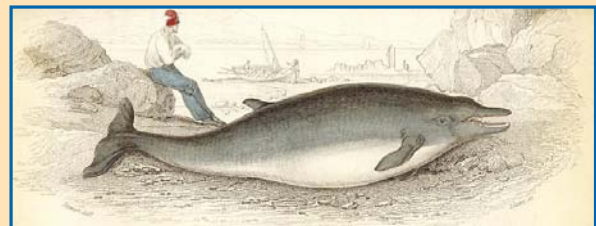
In the mid-1970s, four dead beaked whales were washed ashore separately near San Diego, California. As announced in the latest issue of *Marine Mammal Science* (18, 577–608; 2002), they have now been recognized as a new species. This is something of an event — as with most large mammals, new species descriptions of cetaceans (whales, dolphins and porpoises) are not an everyday occurrence.

Initially, from an examination of the whales' skulls, James Mead tentatively identified the animals as the Southern Hemisphere species *Mesoplodon hectori*. At that time there were a dozen named species of beaked whale (Ziphiidae) within the genus *Mesoplodon*, diagnosed primarily by the size, shape and position of an enlarged pair of teeth in the adult males. An example of a beaked whale — *M. bidens*, drawn in 1843 — is shown in the picture. The enlarged teeth are used in male–male fights, not for feeding; beaked whales feed by using their

tongues in a piston-like manner for sucking in prey, primarily squid.

Merel Dalebout has recently re-examined the remains of the stranded specimens, but this time by analysing DNA-sequence data. She found that data from the California animals clustered far apart from those from *M. hectori*. Dalebout, Mead and colleagues have now formally described the new species, naming it *M. perrini* after William F. Perrin, an eminent contemporary scholar of marine mammals.

The majority of the 85 species of living cetaceans were described from specimens collected during the great voyages of discovery of the eighteenth and nineteenth centuries. Most of those discovered more recently are beaked whales: for example, only 1 out of 34 species of living oceanic dolphins (Delphinidae), but 8 out of 20 beaked whale species, have been described since 1900. The reason is that, as offshore, deep-water creatures,



beaked whales are especially difficult to identify and study.

Dalebout and colleagues' finding is surprising because the Californian coast has long been scoured for stranded specimens, and the offshore waters have been extensively covered during surveys of marine mammal populations. Recognition of a new whale from this region highlights how little we know about biological diversity in the ocean. The discovery also exemplifies the continuing value of specimens collected decades or more ago. To resolve the identity of *M. perrini* with certainty, Dalebout sampled bones of the specimens that had been housed in museum

collections almost a quarter of a century previously.

At 3.9 m long, the adult male specimen of *M. perrini* is smaller than the adult female (4.4 m). This is the reverse of the usual case in mammals, but is characteristic of beaked whales. The high number of calves — three of the five known specimens — is also typical for all species of *Mesoplodon* stranded along the Californian coast, perhaps indicating that they migrate to coastal waters for calving. **John E. Heyning** John E. Heyning is in the Natural History Museum of Los Angeles County, 900 Exposition Boulevard, Los Angeles, California 90007, USA. e-mail: jheyning@nhm.org

modest, and might be expected as a result, rather than a cause, of heart dysfunction.

The authors also engineered 'double-knockout' mice that lack both ACE and ACE2. These animals are identical to the ACE-knockout mice: they have low blood pressure and no heart failure. They presumably make little angiotensin II (because they lack ACE). So Crackower *et al.* again suggest that angiotensin II contributes to the heart-muscle dysfunction seen in the animals lacking just ACE2. The difficulty is that the double-knockout mice also have a much lower blood pressure than the ACE2-deficient animals. Low blood pressure is good for the heart and is often associated with a slower development of heart failure. So it remains to be seen whether the normal heart function in the double-knockout mice results directly from the lack of cardiac angiotensin II or indirectly from the protective effects of low blood pressure.

Finally, Crackower *et al.* studied heart development in fruitflies (*Drosophila melanogaster*). Fruitflies have two ACE2-like enzymes, and the authors looked at mutants that lack one of these, ACER. The mutants die at an early embryonic stage, so the authors analysed the genes expressed by the very earliest heart progenitor cells. They found that two of the genes, named Eve and Tinman, show an abnormal cellular pattern of expression. This implies that ACER is needed for normal heart development in *D. melanogaster*. Exactly how this relates to mammalian biology is unclear. ACE2-deficient mice are born alive; embryonic heart development seems normal and the animals do not develop heart disease until several months after birth. So there seem to be differences between the roles of ACER in flies and ACE2 in mammals — another intriguing issue for further study.

So where are we now in understanding the role of ACE2? Perhaps the most important finding in the paper by Crackower *et al.* is that the mammalian protein affects heart function. Hence, the absence (and perhaps, by extension, the dysfunction) of ACE2 may contribute to heart disease — a killer that claims the lives of nearly one in two people in the developed world. But we still need to know exactly how ACE2 works, and whether and how it helps to control blood pressure in humans. Another question is whether the abnormal heart function seen in the ACE2-knockout mice results from abnormalities in angiotensin peptide levels or whether another, unknown, substrate of ACE2 is affected.

The paper by Crackower *et al.* embodies an important theme in modern angiotensin-related research: an increasing awareness of the many different aspects of physiology that are regulated by the renin-angiotensin system. That ACE2 somehow affects heart-muscle function is now a fact. Investigation of whether this is mediated through

angiotensin II, arguably the central regulator of the human cardiovascular system, or some other mechanism will be the next chapter in the story.

Kenneth E. Bernstein is in the Department of Pathology, Emory University, Atlanta,

Georgia 30322, USA.

e-mail: kbernst@emory.edu

1. Crackower, M. A. *et al.* *Nature* **417**, 822–828 (2002).
2. Tipnis, S. R. *et al.* *J. Biol. Chem.* **275**, 33238–33243 (2000).
3. Donoghue, M. *et al.* *Circ. Res.* **87**, e1–e9 (2000).
4. Kregel, J. H. *et al.* *Nature* **375**, 146–148 (1995).
5. Esther, C. R. Jr *et al.* *Lab. Invest.* **74**, 953–965 (1996).

Semiconductor technology

Imprints offer Moore

R. Fabian Pease

The cost of making chip components smaller using photolithographic printing might soon invalidate Moore's law. A new imprinting technique that can reproduce features as small as 10 nm could save it.

Progress in microelectronics is celebrated in Moore's law, which states that the number of transistors that can be integrated on a single silicon chip doubles every 18 months¹. Moore has also put forward two corollaries to the basic law: that everything gets better as it gets smaller, and that the cost of the manufacturing technology increases geometrically with time².

The semiconductor industry has maintained the validity of Moore's law and of the first corollary by keeping the total chip size about constant and cramming in more, smaller components (transistors and the wires connecting them). The implication of the second corollary is that the increased cost of the manufacturing technology is a threat to the law's continued validity; indeed, if the cost per transistor starts to increase, it could well be the end of the law. But a new imprinting technology for the production of silicon

chips, introduced by Chou *et al.*³ on page 835 of this issue, could keep us on track.

The dominant contributor to the cost of manufacturing chips is the patterning technology, often known as photolithography, in which the image of a chip pattern is optically focused onto a thin film of light-sensitive polymer, or 'resist', on a silicon wafer (Fig. 1a). Selectively dissolving away either the exposed or the unexposed regions of the resist leaves a relief image on the silicon, which can then be transferred, by etching, to the underlying material to make the components that comprise the chip. A sequence of between 4 and 30 overlaid photolithographic levels is needed to make a complete chip.

The key piece of equipment is the optical system, similar in principle to that of a microscope, by which the image is focused onto the resist. Known as a 'stepper', it is complicated

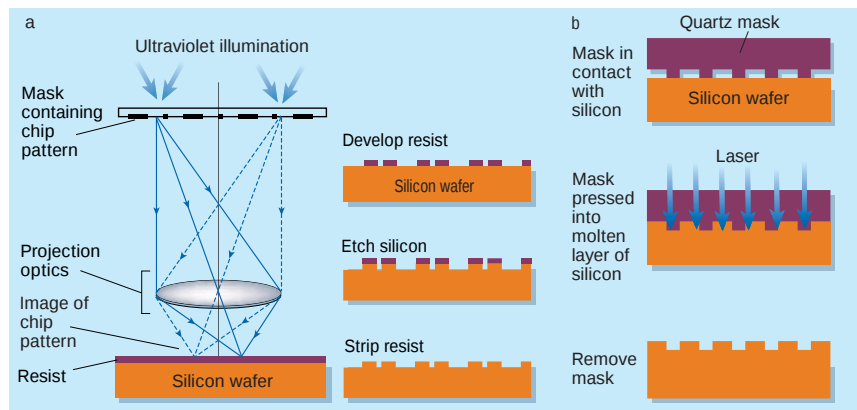


Figure 1 A new recipe for silicon chips? **a**, Currently, electronic components are integrated on a silicon chip using photolithography. The chip pattern is captured on a 'resist' (similar to photographic film) as ultraviolet light is shone through a mask bearing the pattern. After the resist is developed, the silicon layer beneath is etched using the resist pattern as a guide. The resist is removed and the chip pattern in the silicon remains. The future of photolithography, however, may be limited. The projection optics that focus the light onto the resist are costly, and the demand for ever smaller components is unlikely to be met — the smallest features possible at the moment are around 100 nm in size. **b**, Chou *et al.*³ have developed an imprinting technique that they call 'LADI' (laser-assisted direct imprint). Laser light shone through a quartz mask above the silicon surface melts the top layer of silicon. The mask is pressed into the molten layer. When the mask is removed, an accurate imprint of the component design remains in the silicon, with detail rendered at the 10-nm level.

and expensive, currently costing more than US\$10 million. In one minute, a stepper may be required to expose 60 images, each 2×2 cm and containing more than 10 billion features. The smallest possible feature is a square with sides 130 nm long, and the accuracies (or tolerances) of the size and position of features are around 25 nm and 65 nm, respectively; moreover, there must be no detectable defects within the image. To follow Moore's law, the minimum feature size and the respective tolerances will need to be halved over the next three years, while still maintaining the same image size and speed of the process. Over the same period, the stepper cost is projected to rise to more than \$20 million. But with four times as many features created in a single image, the required reduction in cost per exposed feature could still be met.

Physicists will notice that a projected feature size of 65 nm is beyond Rayleigh's resolution limit for a microscope — the minimum distance that can be resolved owing to the effects of diffraction. (This limit is about 110 nm for ultraviolet light with a wavelength of 193 nm.) To achieve this resolution is difficult and expensive, though not impossible, and it is not clear that we can keep reducing feature size and the cost per exposed feature by continuing to push the technique of optical projection. If progress falters, Moore's law will fail and this could be disastrous for the semiconductor industry.

Thus industry is investing in 'next generation lithography', in particular in two favoured technologies, electron-projection and extreme-ultraviolet lithography. Both rely on high-resolution focusing of a projected image. However, progress has been slow and expensive, and there are doubts that either technique can maintain Moore's law.

But a new technology based on conventional, mechanical printing could eliminate the focusing problem altogether. Derived from the imprinting process used to manufacture compact disks, this process can generate sub-micrometre features over an area of 100 cm^2 in less than a second at a cost of about 50 cents — that's between two and three orders of magnitude less than the cost of optical projection. In 1996, Chou and colleagues demonstrated⁴ that a modification of this technique could generate 10-nm features over an image field of about 3 cm^2 , an extraordinary combination of minimum feature size and image size.

There have been other significant advances, such as the development by Whitesides and colleagues⁵ of another mechanical printing process, 'soft lithography', with which features smaller than 100 nm can be printed on non-flat surfaces. In 1998, the US Defense Advanced Research Projects Agency launched a research programme on 'molecular-level, large-area printing' (MLP). And the semiconductor industry is beginning to take notice; results

are emerging from a joint development programme in this area launched by Motorola and the University of Texas⁶.

The latest development reported here by Chou *et al.*³ is one of the most exciting. They have demonstrated that, instead of imprinting in a thin plastic resist film, they can pattern silicon directly by a combination of imprinting and flood illumination (Fig. 1b). They call the process 'LADI', for laser-assisted direct imprint. The printing mask is made of quartz, with a relief image on the surface that presses against the silicon. Flashing laser light through the mask causes the top, sub-micrometre layer of silicon to melt momentarily and take on the shape of the relief image of the mask. The mask is then separated from the silicon, apparently without any unwanted adhesion between the silicon and the quartz.

Chou *et al.* created features of 140-nm dimension (the limit of their mask). But detail at the level of 10 nm on these features was also replicated, indicating the extraordinary resolution of the process. This technique can also be used for patterning polycrystalline films of silicon, one of the most critical steps in patterning silicon chips. Because there is no need for expensive focusing optics, the printing equipment is much simpler than a stepper. And the absence of a resist eliminates that cost too.

But there are technical challenges to be

met. Defects caused and propagated by contact led to the abandoning of contact photo-printing in the early history (pre-1970) of silicon chips, so we might expect similar problems here. But in the decades since, there have been tremendous advances in managing defects. Making a suitable mask with features as small as those required on the wafer is a challenge, but there is at least one tool, Nippon Telephone and Telegraph's experimental electron-beam system 'EBX-3', that seems to be up to the task.

Thus, on grounds of cost, speed and resolution, LADI, or some other form of mechanical printing, may displace optical projection as the preferred manufacturing technology for fashioning silicon chips. So we might expect Moore's law to hold for, maybe, another two decades. ■

R. Fabian Pease is in the Department of Electrical Engineering, Stanford University, Stanford, California 94305, USA.

e-mail: pease@cis.stanford.edu

1. Moore, G. E. *Electronics* **38**, 19 April (1965); <http://www.intel.com/research/silicon/moorespaper.pdf>
2. Moore, G. E. *Wired* **5**, 05 (1997).
3. Chou, S. Y., Keimel, C. & Gu, J. *Nature* **417**, 835–837 (2002).
4. Chou, S. Y., Krauss, P. R. & Renstrom, P. S. *Science* **272**, 85–87 (1996).
5. Xia, Y. & Whitesides, G. *Annu. Rev. Mater. Sci.* **28**, 153–184 (1998).
6. Resnik, D. J. *et al.* in *Emerging Lithographic Technologies VI* (ed. Engelstad, R. L.) *Proc. SPIE* **4688** (in the press).

Neurobiology

Understanding the consequences

Thomas J. Carew

We learn in several ways, one of which involves forming an association between an action and its consequence. Studies of a marine mollusc shed light on how this creature forms a connection between biting and food.

The ability to assess the consequences of one's actions is fundamental to survival: an animal must learn an effective hunting strategy if it is to eat, and to elude predators if it is to live to see another day. Writing in *Science*, Brembs, Lorenzetti and colleagues¹ describe their studies of the neural basis of this type of learning in the marine mollusc *Aplysia*, which could serve as a useful model for understanding more complicated organisms.

This general class of learning is known in the trade as 'operant' or 'instrumental' conditioning². It was first brought into the laboratory over a century ago by Thorndike³, who studied the ability of cats to learn to escape from a 'puzzle box'. At about the same time, another form of learning was discovered by Pavlov⁴, who described how animals form an association between a neutral (also termed a 'conditioned') stimulus, such as a ringing tone, and an 'unconditioned' stimulus that has inherent meaning, such as food. This is known as classical or pavlovian

conditioning. Together these two types of learning provide most of the tools that all animals need to negotiate their environment successfully, enabling them to associate their behaviour with its consequences and to learn predictive relationships about the world.



Figure 1 Learning model — the marine mollusc *Aplysia*. The new paper by Brembs *et al.*¹ gets to grips with how *Aplysia* learns to associate biting with a food reward, a type of 'operant' conditioning.

J. H. BYRNE

In modern neuroscience, the analysis of learning mechanisms is a thriving enterprise. But although classical and operant conditioning are both important, the mechanisms of classical conditioning have received far more attention. Why should this be? The reason lies, at least in part, in how researchers approach the problem. The basic goal of a neurobiological analysis of associative learning is twofold: first, to identify the site of the association in the brain; and second, to characterize the neural mechanisms involved in forming the association. For classical conditioning, this strategy is relatively straightforward. One would first identify the pathways of neurons that respond to the conditioned and unconditioned stimuli. The points at which the two pathways converge would be good candidates for the sites at which an association between the stimuli is formed.

But matters are potentially more complicated in the case of operant conditioning. Here, an association is made not between two stimuli but between an action and its consequence, such as a benefit (or 'reward', in learning parlance) or punishment. So the site of association is not intuitively obvious. For example, it could occur where information about the reward converges with the brain region that initiates the behaviour, which could be more difficult to locate than the regions that respond to conditioned and unconditioned stimuli.

Brembs *et al.*¹ have overcome these difficulties by studying the operant conditioning of feeding in *Aplysia* (Fig. 1). The authors attacked the problem at several levels, from the behaviour of the whole animal down to the electrophysiological properties of single neurons. They focused on feeding in *Aplysia* because this is known to be capable of operant conditioning⁵ and, more importantly, because the neural circuitry underlying feeding has been well characterized.

First, Brembs *et al.* looked at the electrical activity of the oesophageal nerve in whole animals, and found that it increased when the animals ingested food. Presumably, this activity signals the presence of a reward — food. Next, Brembs *et al.* 'trained' the animals to associate spontaneous biting (whether or not food was ingested) with a reward by stimulating the oesophageal nerve, mimicking the usual reward signal, during biting. The result was that the molluscs made significantly more spontaneous bites than controls, both immediately and 24 hours after training. In other words, the operant response (biting) can be reinforced by a food-related reward signal (stimulation of the oesophageal nerve); moreover, the memory of the association between biting and reward can persist for at least 24 hours.

The authors then turned their attention to where this memory might be stored. Here, the previous detailed characterization of the

neural circuitry underlying feeding behaviour was a big help. In the central nervous system, a particular group of neurons — the buccal ganglia — controls biting, and the authors studied one of these, called B51, because it is essential in generating the correct programme of neuronal activity. By recording the electrical activity of B51 neurons in buccal ganglia that had been surgically isolated from *Aplysia*, Brembs *et al.* showed that the burst threshold was lower and the input resistance higher in neurons from trained animals than in controls. The changes in these properties together increase the likelihood that B51 will become active, and hence improve its ability to generate ingestion-related neural programmes.

So it seems that operant conditioning can alter certain properties of the B51 neuron. But it was not clear whether B51 is a genuine site of convergence between bite behaviour and reward — that is, if it is where the association is formed and stored — or whether it is simply affected by that site. To find out, Brembs and colleagues isolated and cultured B51 neurons and examined whether similar changes in properties could be induced by mimicking the operant conditioning procedure at the single-neuron level. They paired the activation of B51 (that would generate a bite) with brief pulses of dopamine, a neuromodulator that probably serves as the reward signal in this system⁶ and many others⁷. They found a significant reduction in burst threshold and increase in input resistance in B51. This did not occur when dopamine application and B51 activation were unpaired.

How do these observations relate to the learned association between biting and reward? The idea is this. B51 activity is required for an animal to attempt to take a bite of food. If that attempt is successful — if food ends up in the mouth — the oesophageal nerve is stimulated. This causes the release of dopamine in the buccal ganglia, increasing the excitability of B51. Given that direct activation of B51 can elicit ingestion-related activity in isolated ganglia⁸, one can envisage that a direct consequence of enhancing B51 excitability would be a greater likelihood of further biting.

So, collectively the data show that B51 is directly affected by operant conditioning, implying that it is at least one important site of memory storage. Moreover, an exciting aspect of the study is that it provides a neuronal mechanism for operant conditioning that can explain not only its associative features, but also how that associative component might control future biting behaviour.

But the significance of the paper extends further. Now that an associative site for operant conditioning has been found, it will be possible to examine the underlying cellular and molecular mechanisms in detail — though this is no easy chore, to be sure. More-



100 YEARS AGO

Mr. Marconi brought forward two interesting pieces of information in his lecture at the Royal Institution last Friday. The first relates to the new form of magnetic detector which he has been employing in place of the coherer. The instrument is found to be more sensitive and trustworthy than the coherer, and gives promise of a great increase in the speed of working. Already a speed of thirty words a minute has been attained, and this may possibly be increased to several hundred. The second point relates to the recent Transatlantic signalling. It seems that on the occasion of Mr. Marconi's journey across the Atlantic in the *Philadelphia*, the signals transmitted during the day failed entirely at a distance of 700 miles, although a message was successfully sent at night more than 1550 miles, and a signal more than 2000 miles. This effect Mr. Marconi suggests may be due to the diselectrification of the aerial waves by the daylight. The difficulty can, however, be got over by the use of greater transmitting power — as is evidenced partly by the fact that the signal received at Newfoundland was transmitted during the daytime.

From *Nature* 19 June 1902.

50 YEARS AGO

It is known that the isotopic composition of carbon in living matter and related materials is different from that in carbonates. About a hundred plants, representing most of the major plant groups, have been investigated... Various hypotheses have been postulated in order to explain the results, and they can easily be described in terms of the 'local carbon dioxide cycle', well known to botanists and geochemists. It is assumed that there is a difference in the rate of assimilation of the light and the heavy carbon dioxide molecules. This difference is accentuated by the cycle: 'local air'—plant—soil—'local air', which works as an isotope enrichment process, assuming that there is an exchange between the 'local' air and the 'main' atmospheric air. At places where this cycle is intense the isotope effect is large; where it is almost absent, for example, in deserts or very windy places, the isotope effect will be small... These results are also of some interest in connexion with the carbon-14 method for age determinations. The observed effects will be accentuated because the difference in rates of assimilation will be much larger.

From *Nature* 21 June 1952.

over, *Aplysia* shows classical conditioning⁹, and much is known about the molecular and cellular mechanisms underlying this form of learning^{10–12}. So it might be possible to compare classical and operant conditioning in *Aplysia* in mechanistic terms. If they have features in common, an exciting principle might emerge: evolution may have come up with a neural 'associative cassette' that can be used in either type of conditioning, depending on the neural circuit in which it is embedded. Of course, this is pure speculation, but the work by Brembs and colleagues will be instrumental in exploring this intriguing possibility. ■

Thomas J. Carew is in the Department of Neurobiology and Behavior, 2205 McGaugh Hall, and the Center for the Neurobiology of Learning and Memory, University of California at Irvine,

California 92697-4550, USA.

e-mail: tcarew@uci.edu

1. Brembs, B., Lorenzetti, F. D., Reyes, F. D., Baxter, D. A. & Byrne, J. H. *Science* **296**, 1706–1709 (2002).
2. Skinner, B. F. *The Behavior of Organisms* (Appleton, New York, 1938).
3. Thorndike, E. L. *Psychol. Rev. Monogr. Suppl.* No. 8 (1898).
4. Pavlov, I. P. *Conditioned Reflexes* (Oxford Univ. Press, Oxford, 1927).
5. Botzner, D., Markovich, S. & Susswein, A. *J. Learn. Mem.* **5**, 204–219 (1998).
6. Nargeot, R., Baxter, D. A., Patterson, G. W. & Byrne, J. H. *J. Neurophysiol.* **81**, 1983–1987 (1999).
7. Schultz, W. *Nature Rev. Neurosci.* **1**, 199–207 (2000).
8. Nargeot, R., Baxter, D. A. & Byrne, J. H. *J. Neurosci.* **19**, 2261–2272 (1999).
9. Carew, T. J., Hawkins, R. D. & Kandel, E. R. *Science* **219**, 397–400 (1983).
10. Hawkins, R. D., Abrams, T. W., Carew, T. J. & Kandel, E. R. *Science* **219**, 400–405 (1983).
11. Walters, E. T. & Byrne, J. H. *Science* **219**, 405–408 (1983).
12. Murphy, G. G. & Glanzman, D. L. *Science* **278**, 467–470 (1997).

Light microscopy

Beyond the diffraction limit?

Ernst H. K. Stelzer

The wave nature of light manifests itself in diffraction, which hampers attempts to determine the location of molecules. Clever use of microscopic techniques might now be circumventing the 'diffraction limit'.

The two best-known physical limitations are Abbe's resolution limit in optical physics and Heisenberg's uncertainty principle in quantum physics. Each defines a natural limit to the resolution or accuracy with which certain parameters can be measured. But, writing in *Physical Review Letters*, Marcus Dyba and Stefan Hell¹ claim to have taken a step beyond one of these limits.

In 1873, Ernst Abbe² realized that the smallest distance that can be resolved between two lines by optical instruments has a physical and not just a technical limit. The distance — the diffraction limit — is proportional to the wavelength and inversely proportional to the angular distribution of the light observed. No matter how perfectly an optical instrument is manufactured, its resolving power will always have this natural limit.

About 50 years later, Werner Heisenberg³ realized that the parameters describing a quantum particle, such as its location and momentum, are not independent. The accuracy with which one parameter can be determined is coupled to the accuracy with which the other can be determined: one can have an accurate idea either of the location or of the momentum, but not of both at the same time. Improvement in accuracy in one parameter will always occur at the expense of decreasing the accuracy in the other. Heisenberg's 'uncertainty principle' is probably one of the most thoroughly tested relationships in physics, a firm foundation of modern quantum theory, and there is no reasonable account that suggests it is incorrect.

Although not a great surprise, it has only recently been shown that Heisenberg's and Abbe's formulae are related⁴.

Over the past few years, several groups have claimed to have achieved image contrasts that have taken the resolution of light microscopes beyond that of classical instruments, and beyond Abbe's limit. For example, confocal fluorescence microscopy has had an enormous impact in biology. Using lasers to induce emission from fluorescent chromophores (the molecules responsible for the colour of the material), it combines point-by-point illumination with synchronous point-by-point detection. This technique has proved especially useful for imaging biological objects because of its optical sectioning capability: deep inside optically dense objects (such as embryos), it is possible to record fluorescent images that show the chromophore distribution in just a single focal plane. The advantage with this type of instrument comes from its nonlinear behaviour — it is essentially sensitive to the square of the light intensity, not to the intensity itself, and this discriminates against light that is out of focus.

On the basis of the work of Lukosz⁵, Gustafsson and colleagues⁶ have achieved excellent resolution with their images of actin protein networks, collected by illuminating the object with light patterns whose intensity varied in a sinusoidal fashion. These and several other contributions to the field have pushed the image resolution down to about 100 nm — which does not contradict Abbe's resolution limit.

Imaging the common soil bacterium *Bacillus megaterium* with focused light of wavelength 760 nm, Dyba and Hell¹ claim to have observed excited molecules with a resolution of 33 nm — that is, down to a distance considerably smaller than Abbe's limit. The authors' technique relies on two essentially independent technologies. The first is '4Pi confocal fluorescence microscopy', which uses two opposing lenses of high numerical aperture to illuminate an object coherently from two sides⁷. This creates an interference pattern that is modulated along the optical axis and reduces the observed volume by a half. The main maximum in the pattern of diffracted light — one 'volume element' in the image sought — is surrounded by other peaks of light intensity, which can generally be removed by further computations and by using knowledge gained about the object with other means.

The second technique is more complicated. First, the fluorophores are excited with a regular, well-focused beam of light. Pico-seconds later, a second beam operating at the emission frequency of the fluorophores but with a light-intensity distribution that has a gap in its centre, stimulates the emission of the fluorophores outside the gap region. A few more picoseconds later, the regular fluorescence emission, stemming from the volume defined by the gap, can be observed. Dyba and Hell used the 4Pi technique to create the gap pattern for the stimulated emission.

The question is what the observed spot width of 33 nm tells us. Have Dyba and Hell achieved a spatial resolution of 33 nm? Although the experiments shown in their paper seem convincing at first, I question whether this is actually a demonstration of resolution improvement. Their images of membranes of *Bacillus megaterium* show a two-dimensional view of well-spaced membranes that are already well resolved in the straightforward confocal image. The images recorded with the new technique seem to localize the two membranes more exactly. This is at best an improvement in the precision of their locations, but not actually in the resolution of the image, which is defined as the minimum distance between two features that can be resolved with a certain contrast⁸.

As already seen in the work of Toraldo di Francia and others^{9,10}, the central diffractive peak can be made much narrower, but this always generates huge side 'lobes' of light, which appear outside the area of interest. This effect is equally visible in the images shown by these authors¹. Extensive computations are required to discriminate the information in the central lobe from that in the side lobes. So the improvement is not entirely due to the instrument but in fact stems from combining the imaging techniques and from using *a priori* knowledge. Another problem is that the signal-to-noise ratio becomes worse.

Nevertheless, although the usefulness of the method introduced by Dyba and Hell is not yet clear, the development is quite fascinating. For the moment one thing is still true: Heisenberg was right and Abbe's limit was certainly not broken. ■

Ernst H. K. Stelzer is in the Cell Biology and Biophysics Programme, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany.

e-mail: stelzer@embl-heidelberg.de

Structural biology

A molecular propeller

Dinshaw J. Patel

Telomeres are protein–DNA structures protecting the ends of chromosomes. The crystal structure of a four-stranded stretch of human telomere DNA, bound to K^+ ions, has implications for the design of anticancer drugs.

The ends of the linear chromosomes of eukaryotes are capped by telomeres — protective structures composed of repetitive DNA bound to proteins. Every time a cell divides it replicates its genetic material, but the replicating enzymes cannot copy the extreme ends of chromosomes. If genes were present right up to the chromosome ends, they would gradually be eroded with each cell division. Telomeres prevent that from happening and, although they are themselves eroded, they can, under some circumstances, be renewed by the enzyme telomerase¹. Among their other functions, telomeres also stop chromosomes from fusing end to end.

The DNA of human telomeres consists of repeats of the nucleotide sequence TTAGGG, ending in a single-stranded segment that overhangs at the end of the double-stranded DNA helix. On page 876 of this issue, Parkinson and colleagues² describe striking crystal structures of single-strand sequences — consisting of two or four TTAGGG repeats — in the presence of K^+ ions, revealing topologies that could readily be incorporated into a higher-order DNA architecture.

In these structures the single-stranded repeats fold up into four-stranded (quadruplex) topologies². The basic building block of any quadruplex is a $G \cdot G \cdot G \cdot G$ tetrad, composed of four hydrogen-bonded guanine nucleotides in a horizontal planar arrangement³ (Fig. 1a). Individual tetrads stack up on each other, with monovalent cations (Na^+ or K^+) sandwiched between them⁴. With TTAGGG repeats, the GGG sequence forms a vertical strand (column) of the quadruplex; the TTA sequence loops over, joining into the GGG sequence of the next repeat, and so on until there are four vertical GGG columns, three horizontal tetrads deep. The GGG columns can adopt either parallel or anti-

1. Dyba, M. & Hell, S. W. *Phys. Rev. Lett.* **88**, 163901 (2002).
2. Abbe, E. *Arch. Mikrosk. Anat.* **9**, 413–468 (1873).
3. Heisenberg, W. Z. *Phys.* **43**, 172–198 (1927).
4. Stelzer, E. H. K. & Grill, S. *Opt. Commun.* **173**, 51–56 (2000).
5. Lukosz, W. *J. Opt. Soc. Am.* **57**, 932–941 (1967).
6. Gustafsson, M. G. L., Agard, D. A. & Sedat, J. W. *J. Microsc.* **195**, 10–16 (1999).
7. Hell, S. & Stelzer, E. H. K. *J. Opt. Soc. Am. A* **9**, 2159–2166 (1992).
8. Stelzer, E. H. K. *J. Microsc.* **189**, 15–24 (1997).
9. Toraldo di Francia, G. *Atti Fond. Giorgio Ronchi* **7**, 366–372 (1952).
10. Martinez-Corral, M., Caballero, M. T., Stelzer, E. H. K. & Swoger, J. *Opt. Express* **10**, 98–103 (2002).

parallel alignments (they are oriented in the same or opposite directions), depending on how the connecting TTA loops are oriented (Fig. 1b–d). Edgewise loops connect adjacent antiparallel strands; diagonal loops connect diagonally opposite antiparallel strands; and double-chain-reversal loops connect adjacent parallel strands.

So how do Parkinson *et al.*'s crystal structures look²? In the four-repeat structure, the GGG columns are all parallel to each other, and the three connecting loops are of the double-chain-reversal type² (Fig. 2a, overleaf). Moreover, all three loops are splayed out like a propeller from the main body of the quadruplex (see Fig. 3b on page 878). They can thus be recognized by proteins that are either components of the

telomere complex or part of the telomere-associated nuclear membrane and matrix.

Parkinson *et al.* also observed the same parallel-stranded alignment in crystals of the two-repeat quadruplex in the presence of K^+ ions. This quadruplex is formed when one two-repeat sequence pairs up with another. It seems that this requires the formation of $A \cdot T \cdot A \cdot T$ tetrad planes, thus expanding the tetrad alphabet beyond $G \cdot G \cdot G \cdot G$ tetrads. Such mixed tetrads might serve a valuable role in cells: they could direct the pairing of homologous chromosomes to enable the chromosomes to 'recombine', swapping segments of DNA.

So the two-repeat structure sheds light on how homologous chromosomes might bind to each other during recombination. The four-repeat structure shows how the basic topology could protect the ends of individual chromosomes. Interestingly, the same four-repeat human telomere sequence adopts a completely different quadruplex architecture in a solution of Na^+ ions⁵ (Fig. 2b). Here, the opposing GGG columns are antiparallel, and there are one diagonal and two edgewise TTA loops. Different monovalent cations can therefore alter the four-repeat quadruplex topology. This might be because K^+ cations (which have an ionic radius of 1.51 Å) are invariably sandwiched between adjacent guanine tetrads⁶, whereas Na^+ cations (ionic radius 1.18 Å) can sometimes be coordinated within a tetrad⁷. Single-nucleotide changes within the telomere repeats might also have a say in cation-mediated folding topology. In mammalian cells, however, telomere repeats are more likely to form K^+ -coordinated quadruplexes as the intracellular K^+ concentration greatly exceeds that of Na^+ (140 mM versus 5–15 mM).

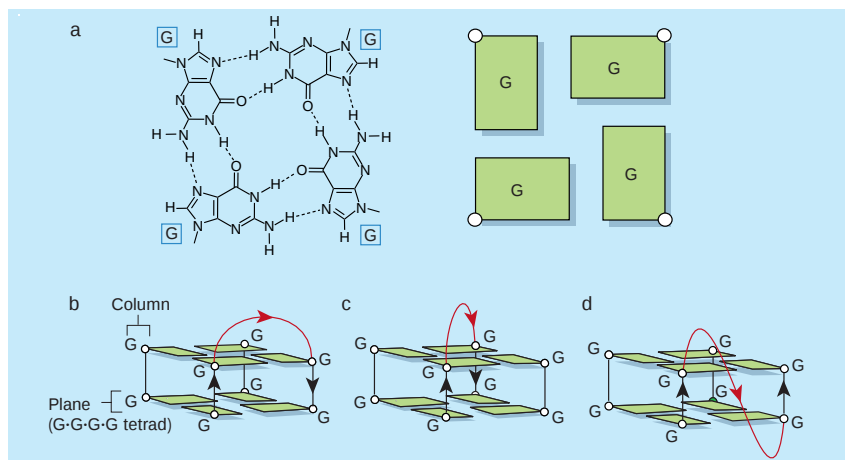


Figure 1 Basics of the quadruplex topology of human telomeric repeat sequences. a, The planar $G \cdot G \cdot G \cdot G$ tetrad alignment, viewed from above, in chemical (left) and schematic (right) form. Left, hydrogen bonds between adjacent guanines are shown by dashed lines. Right, individual guanines are represented as rectangles, and attached sugars as circles. b–d, Stacked tetrads, showing the vertical columns of guanine sequences and horizontal tetrad planes. The DNA backbone of the guanine columns and connecting loops is shown by black and red lines, respectively. Directionalities are shown by arrows. Edgewise loops connect adjacent antiparallel strands; diagonal loops connect opposing antiparallel strands; double-chain-reversal loops connect adjacent parallel strands.

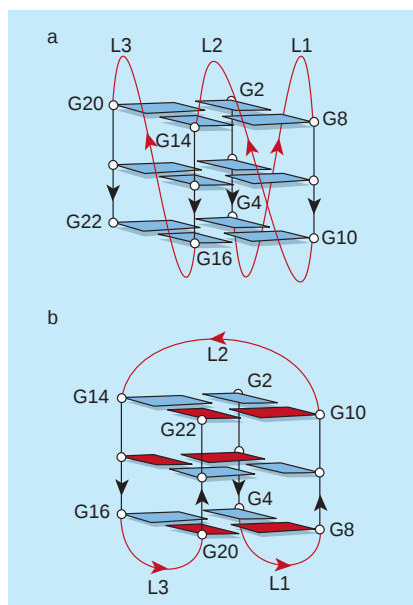


Figure 2 Topology of the quadruplex formed by the four-repeat TTAGGG human telomere DNA sequence $d[AG_3(T_2AG_3)_3]$. **a**, The K^+ -stabilized crystal structure described by Parkinson *et al.*². **b**, The Na^+ -stabilized solution structure⁵. The DNA backbone of the GGG columns and TTA connecting loops is shown by black and red lines, respectively. The guanine residues are shown as rectangles. Guanine can be aligned relative to its sugar in two orientations: *anti* (blue) when it is directed away from its sugar and *syn* (red) when it is positioned over the sugar. In **a**, the columns are all parallel to each other and all three loops (L1–L3) are of the double-chain-reversal type. In **b**, loops L1 and L3 are edgewise type and loop L2 is diagonal, and the opposing columns are antiparallel.

The distinct topologies adopted by the K^+ -coordinated and Na^+ -coordinated four-repeat sequences could have implications for higher-order packing. The diagonal and edgewise loops on the top and bottom of the Na^+ -coordinated structure⁵, together with the fact that the two ends of the sequence are located on the same face, would stop quadruplexes from stacking up on top of each other.

But there is no such barrier to the stacking and subsequent packaging of the compact, disc-like quadruplexes in the K^+ -coordinated structure²: the double-chain-reversal loops are directed outwards in a radial orientation and the chain ends are located on opposite faces of the quadruplex. Double-chain-reversal loops could also facilitate the necessary folding and unfolding of stacked quadruplexes during chromosome replication, without potential complications from the formation of knots in the structure.

Could the quadruplex structure that is described by Parkinson *et al.* help to protect chromosomes from fusing or recombining inappropriately, and stop telomeres from

being mistakenly recognized by the cellular DNA-repair machinery as broken ends? In general terms such events are prevented by telomeres forming a complex with specific proteins⁸, or by insertion of an overhanging single-strand telomeric sequence into an adjacent telomeric double-stranded segment through a process known as 't-loop' formation⁹. It remains to be seen whether quadruplex formation can contribute to this process.

Beyond fundamental chromosome organization, the new results² might also have clinical implications. The enzyme telomerase is needed for complete duplication of telomeric DNA during cell division¹. In humans, it is highly expressed only in tumour cells, enabling them to carry on replicating their chromosomes almost indefinitely. (In non-tumour cells, by contrast, telomerase is not expressed and telomeres gradually erode to the point at which cells cannot duplicate their DNA safely, and so no longer divide.) So telomerase is a promising drug target. It binds to single-stranded telomere ends, and could potentially be inhibited by drugs that compete for these ends (in their quadruplex form). The structure-based design of such drugs¹⁰ necessitates a molecular understanding of the range of quadruplex topologies. So the new structure², together with the previous Na^+ -coordinated structure⁵, is valuable in providing such information.

Finally, the quadruplex scaffold adopted by guanine-rich sequences has been observed in other contexts. For instance, quadruplexes have been associated with genomic regions that control gene transcription, and have been identified *in vivo* by antibody staining in the nuclei of certain protozoans¹¹. Moreover, RNA quadruplexes have been identified in mammalian brain messenger RNAs that are recognized by the fragile X syndrome protein¹². So the architecture of quadruplexes, and their interactions with proteins and other DNA and RNA sequences, should continue to be an active area of research with broad therapeutic implications. ■

Dinshaw J. Patel is in the Cellular Biochemistry and Biophysics Program, Memorial Sloan–Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA.

e-mail: pateld@mskcc.org

1. Blackburn, E. H. *Cell* **106**, 661–673 (2001).
2. Parkinson, G. N., Lee, M. P. & Neidle, S. *Nature* **417**, 876–880 (2002).
3. Gellert, M., Lipsett, M. N. & Davies, D. R. *Proc. Natl Acad. Sci. USA* **48**, 2013–2018 (1962).
4. Sen, D. & Gilbert, W. *Curr. Opin. Struct. Biol.* **1**, 435–438 (1991).
5. Wang, Y. & Patel, D. J. *Structure* **1**, 263–282 (1993).
6. Kang, C. *et al.* *Nature* **356**, 126–131 (1992).
7. Phillips, K. *et al.* *J. Mol. Biol.* **273**, 171–182 (1997).
8. Horvath, M. P. *et al.* *Cell* **95**, 963–974 (1998).
9. Griffith, J. D. *et al.* *Cell* **97**, 503–514 (1999).
10. Hurley, L. H. *Nature Rev. Cancer* **2**, 188–200 (2002).
11. Schaffitzel, C. *et al.* *Proc. Natl Acad. Sci. USA* **98**, 8572–8577 (2001).
12. Darnell, J. C. *et al.* *Cell* **107**, 489–499 (2001).

Daedalus

Cloud-chamber clouds

In the old 'Star Wars' Project, one scheme was to launch particle beams from Earth at the orbiting weaponry in space. The snag, of course, is that such beams are easily absorbed by the atmosphere, especially by clouds. The answer was to launch the beam as a series of pulses. The first beam would travel a certain distance before being absorbed and scattered; the next would ride in the expanded and clarified tunnel driven by the first, and would go further; ultimately the last pulse would break through into clear aerospace and would deliver its deadly energy to the space-borne threat. The whole process would take much less than a second.

Daedalus now connects this idea with the sad fact, bemoaned by every farmer, that endless clouds scud merrily over his fields but deposit no rain on his parched land. Many attempts have been made to make such annoying clouds rain, by seeding them from aircraft or even by firing cannon at them from the ground. Daedalus's scheme recalls the latter. He plans to fire a carbon dioxide laser at the clouds, followed by an electron beam or one of alpha particles. Unlike the Star Wars scheme, these co-linear beams will not reach space. They are intended to be broken up in the clouds, ideally near the tops of the most promising ones.

Wilson's cloud chamber is the crucial experiment here. In this device, the tiny racing fundamental particles precipitate enormously larger droplets along their track, revealing the particles to watchers or cameras. Daedalus reckons that the disruption of a cloud top by his beam system should release enormous numbers of wilsonian droplets, which will fall slowly through the cloud. As in ordinary rainfall, they will be further amplified by collision. Each droplet will collide with those beneath. It will grow in the process, fall faster, and collide more frequently with more cloud droplets. Ultimately, full-sized raindrops will fall out of the base of the cloud.

The beam gadget, though smaller and cheaper than the Star Wars device, will still be costly. Farmers will probably hire it by the day, and use it to wet all their land for that day. But it should then transform farming. It should make it a far more predictable business, less dependent on chance water. A cloud is just a mass of small water droplets: it ought to rain, and any device that encourages this process is going with the grain of atmospheric physics.

David Jones

Only male fin whales sing loud songs

These mammals need to call long-distance when it comes to attracting females.

The low-frequency vocalizations of fin and blue whales are the most powerful and ubiquitous biological sounds in the ocean^{1,2}. Here we combine acoustic localization and molecular techniques to show that, in fin whales, only males produce these vocalizations. This finding indicates that they may function as male breeding displays, and will help to focus concern on the impact of human-generated low-frequency sounds on recovering whale populations.

The long, patterned 15–30-Hz ('20-Hz') vocal sequences of fin whales (*Balaenoptera physalus*) can reach intensities of 184–186 decibels (dB) relative to 1 μ Pa of sound pressure, and can be detected throughout the world's oceans^{3,4}. However, the source of these vocalizations was hard to identify⁵ because of the inherent difficulty of studying wide-ranging pelagic species and in locating low-frequency sound sources in the ocean. Until now, their function was therefore unknown⁶, making it difficult to assess the effects of increasing levels of human-produced sounds on these whales⁷.

We investigated whether these vocalizations could be breeding displays⁸ by determining the sex of vocalizing fin whales (Fig. 1) and comparing the results to the overall sex ratio in Loreto Bay, Gulf of California, Mexico. We identified and tracked vocalizing fin whales by using a 120-metre-long towed array of hydrophones and beam-forming software to compute a series of cross-bearings to a vocalizing individual. Required confirmation criteria included the absence of other whales that could potentially match the acoustic cross-bearings, and asynchronization of vocalizations with respiration.

Once the vocalizing animal was unambiguously identified, a skiff was deployed to obtain a biopsy sample in order to ascertain its sex⁹. To estimate the overall sex ratio in the study area, we took biopsy samples from all fin whales encountered during systematic weekly surveys.

Vocalizations were produced only by male fin whales, despite a 1:1 overall sex ratio in the area (vocalizing: males, 9; females, 0; binomial test, $P < 0.001$; population: males, 21; females, 22; binomial test with normal approximation, $Z_{0.05,2} = 0.153$, $P = 0.879$). This sexual dichotomy in vocal behaviour supports the idea that the patterned sounds of fin whales are male breeding displays.

We propose that these displays serve to attract females from great distances to aggregations of patchily distributed prey. This is supported by several observations.



Figure 1 Only boys make noise: the wide dispersal of populations of the fin whale, seen here surfacing off the Mexican coast, means that males have to sing extra loudly to woo females.

First, fin and blue whales (*Balaenoptera musculus*) do not aggregate in specific areas for breeding, a behavioural trait that distinguishes them from the closely related humpback whale (*Megaptera novaeangliae*)¹⁰. Second, we previously found that fin whales use the Loreto study area to forage on dense aggregations of krill¹¹. Third, the low-frequency vocalizations of *Balaenoptera* spp. are optimal for long-distance communication in deep water¹².

Our results help to focus growing concern over the effects of human-produced sound on *Balaenoptera* spp.¹³. Sound levels from commercial ships, military sonar, seismic surveys and ocean acoustic research are extremely high (190–250 dB relative to 1 μ Pa at 1 m; ref. 4) and, at least since the early 1960s, the amount of human-produced sound in the frequency range used by large whales has increased⁷. A sound is detectable if its received level exceeds that of background noise by enough to be detected by the animal. An increase in ambient noise could thus reduce the distance over which receptive females might hear the vocalizations of males. To the extent that growth of *Balaenoptera* populations is limited by

the encounter rate of receptive females with singing males, the recovery of fin- and blue-whale populations from past exploitation could be impeded by low-frequency sounds generated by human activity.

Donald A. Croll*, Christopher W. Clark†, Alejandro Acevedo‡, Bernie Tershy*, Sergio Flores§, Jason Gedamke*, Jorge Urban§

*Department of Ecology and Evolutionary Biology, Center for Ocean Health, University of California, Santa Cruz, California 95060, USA

e-mail: croll@biology.ucsc.edu

†Cornell Laboratory of Ornithology, Bioacoustics Research Program, 159 Sapsucker Woods Road, Ithaca, New York 14850, USA

‡California Academy of Sciences, 55 Concourse Drive, Golden Gate Park, San Francisco, California 94118, USA

§Universidad Autónoma de Baja California Sur, AP 19-B, La Paz, BCS 23080, Mexico

1. Aroyan, J. L. et al. in *Hearing by Whales and Dolphins* (eds Au, W. W. L., Popper, A. N. & Fay, R. N.) 409–469 (Springer, New York, 2000).
2. Thode, A. M., D'Spain, G. L. & Kuperman, W. A. *J. Acoust. Soc. Am.* **107**, 1286–1300 (2000).
3. Charif, R. A., Mellinger, D. K., Dunsmore, K. J., Frstrup, K. M. & Clark, C. W. *Mar. Mammal Sci.* **18**, 81–98 (2002).
4. Richardson, W. J., Greene, C. R., Malm, C. I. & Thomson, D. H. *Marine Mammals and Noise* (Academic, New York, 1995).
5. Schevill, W. E., Watkins, W. A. & Backus, R. H. in *Marine Bio-Acoustics* (ed. Tavolga, W. N.) 147–152 (Pergamon, New York, 1964).
6. Popper, A. N., Hawkins, H. L. & Gisiner, R. C. *Bioacoustics* **8**, 163–182 (1997).
7. Andrew, R. K., Howe, B. M. & Mercer, J. A. *Acoust. Res. Lett. Online* **3**, 65–70 (2002).
8. Watkins, W. A., Tyack, P., Moore, K. E. & Bird, J. E. *J. Acoust. Soc. Am.* **82**, 1901–1912 (1987).
9. Berube, M. & Palsboll, P. *Mol. Ecol.* **5**, 283–287 (1996).
10. Clapham, P. J. *Mammal Rev.* **26**, 27–49 (1996).
11. Croll, D. A., Acevedo-Gutierrez, A., Tershy, B. R. & Urban-Ramirez, J. *Comp. Biochem. Physiol. A* **129**, 797–809 (2001).
12. Clark, C. W. & Ellison, W. T. in *Echolocation in Bats and Dolphins* (eds Thomas, J., Moss, C. & Vater, M.) (Univ. Chicago Press, in the press).
13. Payne, R. S. & Webb, D. *Ann. NY Acad. Sci.* **188**, 110–141 (1971).

Competing financial interests: declared none.

Physiology

Dynamic instabilities in the inflating lung

In lung diseases such as asthma¹, expiratory flow becomes limited², airways can collapse³ and the vital exchange of gases is compromised. Here we model the inflation of collapsed regions of the lung during inspiration in terms of avalanches propagating through a bifurcating network of airways, and find that the accompanying cascade of dynamic pressure instabilities — avalanche 'shocks' — manifests as negative elastic resistance of the lung. Our analysis

of this apparent thermodynamic paradox provides a better understanding of aeration in the deep regions of the lung, which may find application in medical conditions in which gas exchange is impaired.

Most thermodynamic systems respond to an applied load by developing a restoring force — for example, when air is pumped into a rigid chamber, the pressure inside the chamber increases monotonically. However, during inflation of the mammalian lung from the collapsed state, the lung does not always develop an increasing restoring force; instead, the pressure inside the lung decreases intermittently. We explain this apparently paradoxical behaviour in terms

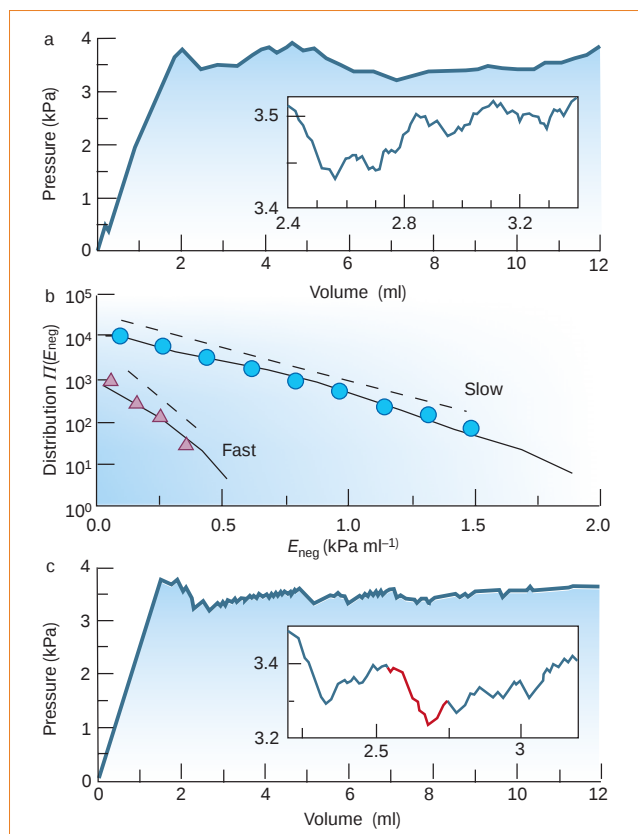


Figure 1 Pressure–volume curves and distributions of negative elastance. **a**, Example of the P – V curve (see text) during the inflation of a degassed rat lung. Inset, magnification of a region with many local negative-elastance patterns. **b**, Distributions of negative elastance from 10 independent inflations at rates of 2.0 ml s⁻¹ (triangles) and 0.5 ml s⁻¹ (circles). The values of E_0 obtained from the straight-line fits to the measured distributions are 0.28 kPa ml⁻¹ and 0.1 kPa ml⁻¹ for the slow and fast inflations, respectively (dashed lines). Solid lines correspond to the distributions of negative elastance from 1,000 simulated inflations of an 18-generation symmetric binary tree. **c**, Example of the P – V curve from inflation of the model. Inset, magnification of a region with many local negative-elastance patterns similar to those in **a**; red line, trace of an avalanche shock.

of a non-equilibrium avalanche model.

We inflated isolated, degassed rat lungs to total lung capacity at two different but constant rates by pumping in 12 ml of air with a computer-controlled piston. The lung recoil pressure, P , was then measured with respect to atmosphere as a function of the volume displacement, V , of the piston. With increasing V , P also increases, but this increase is not monotonic and P intermittently decreases (Fig. 1a), causing the elastance, E (defined as $E = dP/dV$), to take negative and positive values intermittently.

We estimate E as the local slope of a straight-line fit within a moving window along the P – V curve. The magnitude, E_{neg} , of elastances with negative values fluctuates. We find that the distribution $\Pi(E_{\text{neg}})$ is exponential,

$$\Pi(E_{\text{neg}}) \propto e^{-(E_{\text{neg}}/E_0)} \quad (1)$$

where E_0 , the characteristic value of E_{neg} , increases with inflation rate (Fig. 1b). The pattern of E_{neg} varies from inflation to inflation, but the distribution in equation (1) is reproducible. (This excludes the possibility that E_{neg} arises from tissue rupture.)

To understand the existence and fluctuations of the negative elastance, we developed a dynamical model of a system consisting of a piston chamber and a lung. We modelled the lung as a binary tree terminating in elastic alveoli⁴. The root of the tree, the trachea, is connected to the

chamber containing 12 ml of air at atmospheric pressure, corresponding to the experimental conditions. Inflation is simulated by increasing V at a constant rate.

At the start of the inflation, all airways are closed and no alveoli are connected to the root, so a small increase in V increases P . A closed airway opens when P exceeds the airway's opening threshold pressure⁵, P_{th} , which we assume to be a random variable uniformly distributed between 0 and 4 kilopascals (the pressure at total lung capacity). When a segment opens, the pressure propagates deeper into the airway tree. If the P_{th} of either daughter airway is also smaller than P , then this daughter opens together with its parent.

This process leads to an avalanche of openings involving a large number of airways and alveoli⁶. The newly opened airways and alveoli increase the volume of the lung, and P decreases according to Boyle's law. However, the reduction in P can terminate the propagation of the avalanche and P increases again owing to the steady influx of air — we term this phenomenon an avalanche shock (Fig. 1c, inset), which is not a conventional propagating shock wave.

The time course of an avalanche shock is smoothed by relaxation processes due to flow resistance in the airways⁷ and the viscoelasticity of the alveolar walls⁸. The decreasing component of an avalanche shock corresponds to instabilities that are characterized by negative elastance. The

continuous increase in P is intermittently interrupted by avalanche shocks of different magnitudes until all of the air in the chamber is injected into the lung.

The model produces P – V curves (Fig. 1c) that are similar to those observed experimentally (Fig. 1a). The local E_{neg} is determined in the same way as for the measured data. The distributions $\Pi(E_{\text{neg}})$ from the simulations fit the experimentally obtained distributions for both slow and fast inflation rates (Fig. 1b). These results indicate that inflating the lung from a state in which a considerable part of the gas-exchange region is collapsed is a non-equilibrium dynamical process characterized by a sequence of instabilities with negative elastance.

For slow inflation, the system has enough time after each avalanche to reach equilibrium, so the individual avalanche shocks are distinct and non-overlapping. With increasing inflation rate, however, avalanche shocks due to separate avalanches increasingly overlap, resulting in smoother P – V curves with fewer regions of negative elastance. We conclude that the paradoxical negative elastance and its distribution arise from avalanche shocks involving the sudden recruitment and subsequent relaxation of a large number of airways and alveoli.

Instabilities related to airway and alveolar collapse and recruitment have a key function in impaired gas exchange in premature infants⁹, in sufferers of severe asthma³ or after acute lung injury¹⁰. These instabilities have previously been attributed to a combination of the static properties of the alveolar liquid lining¹¹ and the non-uniform deformation of lung tissue¹². Our results and quantitative modelling provide a new interpretation of these instabilities.

Adriano M. Alencar*†, Stephen P. Arold*, Sergey V. Buldyrev†, Arnab Majumdar†, Dimitrije Stamenović*, H. Eugene Stanley†, Béla Suki*

*Department of Biomedical Engineering, and

†Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA

e-mail: adriano@bu.edu

1. Tobin, M. J. *Am. J. Crit. Care Med.* **165**, 598–618 (2002).
2. Mead, J., Turner, J. M., Macklem, P. T. & Little, J. B. *J. Appl. Physiol.* **22**, 95–108 (1967).
3. Lutchen, K. R., Kaczka, D. W., Israel, E., Suki, B. & Ingenito, E. P. *Am. J. Crit. Care Med.* **164**, 207–215 (2001).
4. Shlesinger, M. F. & West, B. J. *Phys. Rev. Lett.* **67**, 2106–2108 (1991).
5. Gaver, D. P., Samsel, R. W. & Solway, J. J. *J. Appl. Physiol.* **69**, 74–85 (1990).
6. Suki, B. et al. *Nature* **368**, 615–618 (1994).
7. Pedley, T. J., Schroter, R. C. & Sudlow, M. F. *Respir. Physiol.* **9**, 387–405 (1970).
8. Fukaya, H., Martin, C. J., Young, A. C. & Katsura, S. *J. Appl. Physiol.* **25**, 689–695 (1968).
9. Pattle, R. E., Claireaux, A. E., Davis, P. A. & Cameron, A. H. *Lancet* **2**, 469–473 (1962).
10. Muscedere, J. G., Mullen, J. B., Gan, K. & Slutsky, A. S.

Am. J. Crit. Care Med. **149**, 1327–1334 (1994).
 11. Clements, J. A., Hustead, R. F., Johnson, R. T. & Gribetz, I. J. Appl. Physiol. **16**, 444–450 (1961).
 12. Stamenovic, D. & Wilson, T. A. J. Appl. Physiol. **73**, 596–602 (1992).
 Competing financial interests: declared none.

Surface phenomena

Contact time of a bouncing drop

When a liquid drop lands on a solid surface without wetting it, it bounces with remarkable elasticity^{1–3}. Here we measure how long the drop remains in contact with the solid during the shock, a problem that was considered by Hertz⁴ for a bouncing ball. Our findings could help to quantify the efficiency of water-repellent surfaces (super-hydrophobic solids⁵) and to improve water-cooling of hot solids, which is limited by the rebounding of drops⁶ as well as by temperature effects.

The way in which a water drop of radius R deforms during its impact with a highly hydrophobic solid depends mainly on its impinging velocity, V . The Weber number, $W = \rho V^2 R / \gamma$, compares the kinetic and surface energies of the drop, where ρ and γ are the liquid density and surface tension, respectively. The greater the value of W , the larger are the deformations that occur during the impact (Fig. 1).

High-speed photography (Fig. 1) enabled us to measure the drop's contact time, τ . The frame rate could be greater than 10^4 Hz,

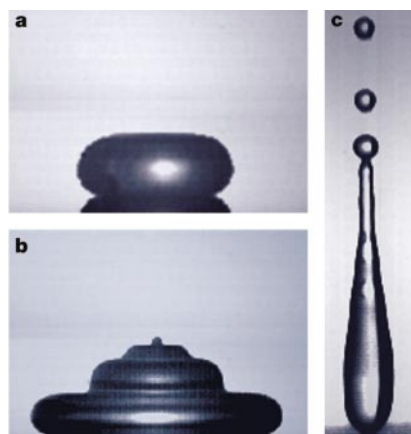


Figure 1 Millimetre-sized water drops with different Weber numbers (W) hitting a super-hydrophobic solid. W compares the kinetic and surface energies of the drop ($W = \rho V^2 R / \gamma$, where R is the drop radius, V is the impact velocity, and ρ and γ are the density and surface tension, respectively, of the liquid). **a**, When W is close to unity, the maximum deformation during contact becomes significant. **b**, When $W \approx 4$, waves develop along the surface and structure the drop. **c**, When $W \approx 18$, the drop becomes highly elongated before detaching and gives rise to droplets; however, the contact time is independent of the details of the impact (see Fig. 2a).

allowing precise measurements of τ , which we found to be in the range 1–10 ms. As the impact is mainly inertial (with a restitution coefficient² as great as 0.91), τ is expected to be a function of only R , V , ρ and γ , and thus to vary as $R/V.f(W)$. For a Hertz shock, for example, the maximum vertical deformation, δ , scales as $R(\rho V^4/E^2)^{1/5}$, where E is the Young's modulus of the ball⁷. Taking a drop's Laplace pressure, $E \approx \gamma/R$, as an equivalent modulus and noting that $\tau \approx \delta/V$, we find for a Hertz drop that $f(W) \sim W^{2/5}$ and that the contact time varies as $V^{-1/5}$ and $R^{7/5}$.

Figure 2a shows that the contact time does not depend on the impact velocity over a wide range of velocities (20–230 cm s^{-1}), although both the deformation amplitude and the details of the intermediate stages largely depend on it. This is similar to the case of a harmonic spring, although oscillations in the drop are far from being linear. Moreover, this finding confirms that viscosity is not important here.

Figure 2b shows that τ is mainly fixed by the drop radius, because it is well fitted by $R^{3/2}$ over a wide range of radii (0.1–4.0 mm). Both this result and the finding shown in Fig. 2a can be understood simply by balancing inertia (of the order $\rho R/\tau^2$) with capillarity (γ/R^2), which yields $\tau \approx (\rho R^3/\gamma)^{1/2}$, of the form already stated with $f(W) \sim W^{1/2}$. This time is slightly different from the Hertz time because the kinetic energy for a solid is stored during the impact in a localized region, whereas in our case it forces an overall deformation of the drop (Fig. 1).

The scaling for τ is the same as for the period of vibration of a drop derived by Rayleigh⁸, and is consistent with a previous postulation⁹, although the motion here is asymmetric in time, forced against a solid, and of very large amplitude. Absolute

values are indeed found to be different: the prefactor deduced from Fig. 2b is 2.6 ± 0.1 , which is significantly greater than $\pi/\sqrt{2} \approx 2.2$ for an oscillating drop⁸. Another difference between the two systems is the behaviour in the linear regime ($W \ll 1$): for speeds less than those shown in Fig. 2, we found that τ depends on V , and typically doubles when V is reduced from 20 to 5 cm s^{-1} , which could be due to the drop's weight¹⁰.

The brevity of the contact means that a drop that contains surfactants, which will spread when gently deposited onto the solid, can bounce when thrown onto it; this is because the contact time is too short to allow the adsorption of the surfactants onto the fresh interface generated by the shock. Conversely, the contact time should provide a measurement of the dynamic surface tension of the drop.

Denis Richard*, Christophe Clanet†, David Quéré*

*Laboratoire de Physique de la Matière Condensée, URA 792 du CNRS, Collège de France, 75231 Paris Cedex 05, France

e-mail: quere@ext.jussieu.fr

†Institut de Recherche sur les Phénomènes Hors Équilibre, UMR 6594 du CNRS, BP 146, 13384 Marseille Cedex, France

- Hartley, G. S. & Brunsell, R. T. in *Surface Phenomena in Chemistry and Biology* (ed. Danielli, J. F.) 214 (Pergamon, Oxford, 1958).
- Richard, D. & Quéré, D. *Europhys. Lett.* **50**, 769–775 (2000).
- Aussillous, P. & Quéré, D. *Nature* **411**, 924–927 (2001).
- Hertz, H. J. *Reine Angew. Math.* **92**, 156–171 (1881).
- Nakajima, A., Hashimoto, K. & Watanabe, T. *Monatsh. Chem.* **132**, 31–41 (2001).
- Frohn, A. & Roth, R. *Dynamics of Droplets* (Springer, Berlin, 2000).
- Landau, L. D. & Lifschitz, E. M. *Theory of Elasticity* 3rd edn (Pergamon, Oxford, 1986).
- Rayleigh, Lord *Proc. R. Soc. Lond. A* **29**, 71–97 (1879).
- Wachters, L. H. J. & Westerling, N. A. J. *Chem. Eng. Sci.* **21**, 1047–1056 (1966).
- Perez, M. et al. *Europhys. Lett.* **47**, 189–195 (1999).

Competing financial interests: declared none.

Evolutionary biology

Hedgehog crosses the snail's midline

According to the dorsoventral axis-inversion theory¹, protostomes (such as insects, snails and worms) are organized upside-down by comparison with deuterostomes (vertebrates)^{2–5}, in which case their respective ventrally (belly-side) and dorsally (back-side) located nervous systems, as well as their midline regions, should all be derived from a common ancestor⁵. Here we provide experimental evidence for such homology by showing that an orthologue of *hedgehog*, an important gene in midline patterning in vertebrates, is expressed along the belly of the larva of the limpet *Patella vulgata*. This

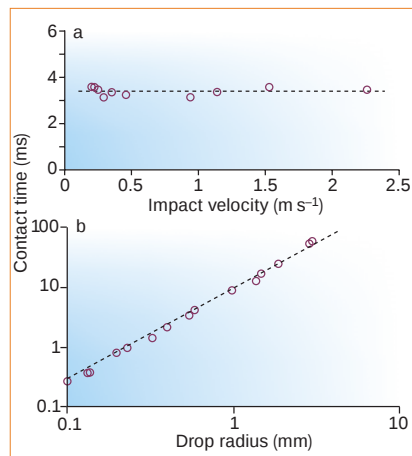


Figure 2 Contact time of a bouncing drop as a function of impact velocity and drop radius. **a**, **b**, In the explored interval (Weber number, W , between 0.3 and 37), the contact time is **a**, independent of the impact velocity, V , but **b**, depends on the drop radius, R . Dotted lines indicate slopes of 0 (**a**) and $3/2$ (**b**).

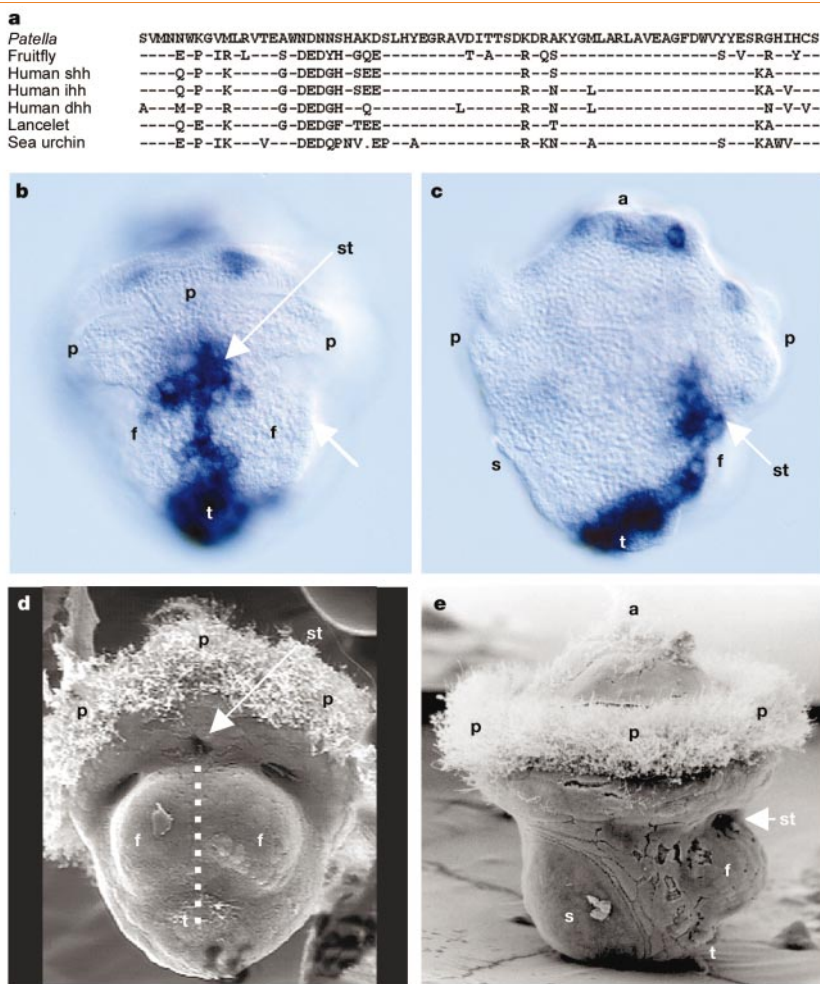


Figure 1 The *hedgehog* gene in the limpet *Patella vulgata* is expressed in the ventral midline of the trochophore larva. **a**, Alignment of part of the deduced *hedgehog* amino-acid sequence from *P. vulgata* with *hedgehog* proteins from fruitfly (*Drosophila*), human (*Homo*), *sonic*, *indian* and *desert* variants, lancelet (*Branchiostoma*) and sea urchin (*Lytechinus*). Dashes denote residues identical to those at the same position in the *hedgehog* protein of *Patella*. (The *hedgehog* gene sequence of *Patella* has been deposited at GenBank under accession no. AF435840). **b**, **c**, Expression of the *hedgehog* gene of *Patella*, as revealed by *in situ* hybridization (specific stain for *hedgehog* messenger RNA; dark blue), in whole-mount embryos (24-h-old pretorsional trochophore larvae) shown from the ventral side (**b**) and the right side (**c**). **d**, **e**, Scanning electron micrographs of larvae of the same age, shown from the ventral side (**d**) and the right side (**e**). Dashed line in **d** shows the ventral midline running from the future mouth (stomodaeum), over the foot and down towards the ciliated telotroch. Anterior is at the top. a, apical tuft; s, developing shell; st, stomodaeum; p, prototroch; f, foot; t, telotroch.

finding supports the existence of a similar mechanism for the development of the midline of the nervous system in protostomes and deuterostomes.

In protostomes as well as deuterostomes, the midline is an embryonic region that functions in patterning of the adjacent nervous tissue^{6,7}. The ventral midline in insects is the region from which cells detach to form the ventrally located nerve cords; in vertebrates, the midline is originally located dorsally. During development, it folds inwards and becomes the ventral part of the dorsally located neural tube and is then called the ventral midline, or floor plate.

The dorsoventral axis-inversion theory predicts that similar genes should participate in embryonic patterning of the midline region in both protostomes and deuterostomes,

but so far this has not been demonstrated⁵. *Sonic hedgehog* is expressed in the midline of the early vertebrate embryo, most prominently in the floor plate of the developing nervous system, where it functions in the differentiation of ventral neural-tube structures and the dorsoventral patterning of adjacent tissues⁶. However, in the protostome insect *Drosophila* it is not expressed in the ventral midline — instead, it sets up the border between segments⁸, a finding that argues against homology of the deuterostome and protostome midlines.

brief communications is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on *Nature's* website (www.nature.com/nature) or on request from nature@nature.com

We isolated a *hedgehog* homologue (the predicted partial amino-acid sequence of which is shown in Fig. 1a) from the snail *Patella vulgata* and used it as a probe for *in situ* hybridization of whole larvae to determine the site of its expression. We obtained a positive hybridization signal in the ectodermal cells of the ventral midline of a 24-hour-old snail larva (Fig. 1b, c) — the ventral midline is evident as the region running from what will become the mouth, over the foot, and down to a ciliated structure below the foot known as the telotroch (Fig. 1d, e).

To our knowledge, this expression of *hedgehog* in *Patella* is the first example of an orthologue (gene homologue) of a deuterostome midline gene being expressed in the ventral midline of a protostome larva. Although the *Patella* trochophore larva does not contain a ciliated band at the ventral midline, other larvae of this type do^{9–11}. There is good evidence that the nervous systems of deuterostomes and protostomes are both derived from ciliated bands (see refs 12, 13, for example), so the *hedgehog* gene of *Patella* could be involved in nervous-system development, supporting the idea of a homologous mechanism for midline development in protostomes and deuterostomes.

Further similarities in nervous-system development between protostomes and deuterostomes may come to light by looking at other understudied organisms besides *P. vulgata*. The striking correspondence in midline signalling between vertebrates and this snail points to an evolutionarily ancient role of the *hedgehog* gene in early patterning of the nervous system.

Alexander J. Nederbragt,

André E. van Loon, Wim J. A. G. Dictus

Department of Developmental Biology,
Utrecht University, Padualaan 8, 3584 CH Utrecht,
The Netherlands

e-mail: alexanderj@nederbragt.net

1. Arendt, D. & Nübler-Jung, K. *Nature* **371**, 26 (1994).
2. Geoffroy St Hilaire, E. *Mém. Mus. Hist. Nat.* **9**, 89–119 (1822).
3. De Robertis, E. M. & Sasai, Y. *Nature* **380**, 37–40 (1996).
4. Arendt, D. & Nübler-Jung, K. *Mech. Dev.* **61**, 7–21 (1997).
5. Arendt, D. & Nübler-Jung, K. *Development* **126**, 2309–2325 (1999).
6. Sasai, Y. & De Robertis, E. M. *Dev. Biol.* **182**, 5–20 (1997).
7. Menne, T. V., Luer, K., Technau, G. M. & Klammt, C. *Development* **124**, 4949–4958 (1997).
8. Lawrence, P. A. & Struhl, G. *Cell* **85**, 951–961 (1996).
9. von Werner, B. *Helgol. Wiss. Meeresunters.* **5**, 169–217 (1955).
10. Morril, J. B. in *Developmental Biology of Freshwater Invertebrates* 399–483 (Liss, New York, 1982).
11. Kennedy, J. J. & Keegan, B. F. *J. Mar. Biol. Assoc. UK* **72**, 783–805 (1992).
12. Crowther, R. J. & Whittaker, J. R. *J. Neurobiol.* **23**, 280–292 (1992).
13. Nielsen, C. *Animal Evolution* (Oxford Univ. Press, Oxford, 2001).

Competing financial interests: declared none.

Ordered porous materials for emerging applications

Mark E. Davis

Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA

“Space—the final frontier.” This preamble to a well-known television series captures the challenge encountered not only in space travel adventures, but also in the field of porous materials, which aims to control the size, shape and uniformity of the porous space and the atoms and molecules that define it. The past decade has seen significant advances in the ability to fabricate new porous solids with ordered structures from a wide range of different materials. This has resulted in materials with unusual properties and broadened their application range beyond the traditional use as catalysts and adsorbents. In fact, porous materials now seem set to contribute to developments in areas ranging from microelectronics to medical diagnosis.

Porous solids are of scientific and technological interest because of their ability to interact with atoms, ions and molecules not only at their surfaces, but throughout the bulk of the material. Not surprisingly, traditional applications of porous materials thus involve ion exchange, adsorption (for separation) and catalysis, and many of these benefit from the high order that can be achieved in solids such as zeolites.

The pores of solids are classified according to size: pore sizes in the range of 2 nm and below are called micropores, those in the range of 2 nm to 50 nm are denoted mesopores, and those above 50 nm are macropores. The distribution of sizes, shapes and volumes of the void spaces in porous materials directly relates to their ability to perform the desired function in a particular application. The need to create uniformity within the pore size, shape and volume has steadily increased over recent years because it can lead to superior applications properties. For example, a material with uniform micropores, such as a zeolite, can separate molecules on the basis of their size by selectively adsorbing a small molecule from a mixture containing molecules too large to enter its pores. Clearly, a distribution of pore sizes would limit the ability of the solid to separate molecules of differing sizes.

In addition to the pore space, the atoms in the solid creating that space can be important. For example, molecular sieves comprising pure silica are hydrophobic and can adsorb organic components from water, whereas molecular sieves comprising aluminosilicates are hydrophilic and can thus adsorb water from organic solvents. Because the control over the uniformity of pore space and the composition of the solid that creates the space is of importance, this review concentrates on porous materials with more ordered structures. Here, I will cover significant issues in microporous materials that have occurred over the last decade or so. In addition to microporous materials, the well-ordered mesoporous solids will be discussed, as they have recently received much attention. The type of materials described here are generally synthesized at low temperatures (less than 473 K) and pressures (a few atmospheres) and in aqueous media. The syntheses do not normally give the thermodynamically most stable product but rather a material that is the result of a kinetically controlled synthetic pathway. Although the starting reagents and compositions and the precise procedures will differ for each type of material, the syntheses are reproducible and thus allow for use in applications. A brief review on zeolites and molecular sieve syntheses that discusses these issues is available¹, as are reviews on microporous and ordered mesoporous materials syntheses and their applications as catalysts².

In 1988, the first report of a crystalline microporous material with uniform pores larger than 1.0 nm appeared³, the ability to syn-

thesize well-ordered mesoporous materials was about to be announced⁴, and scaffolding-like (but non-porous) structures built from three-dimensional, linked modular molecular units⁵ were appearing in the literature. Moreover, zeolite-based membranes with ‘molecular sieving’ properties were discussed⁶, and the inclusion of electro- and photo-active guest molecules in porous host materials⁷ gave rise to interesting phenomena. But despite their promise for advanced materials applications, none of these materials led to practical successes. Today, the situation is rather different: our understanding of the structure of ordered porous materials and how to control and tune it has increased considerably and advances have led to practical applications. In fact, the future

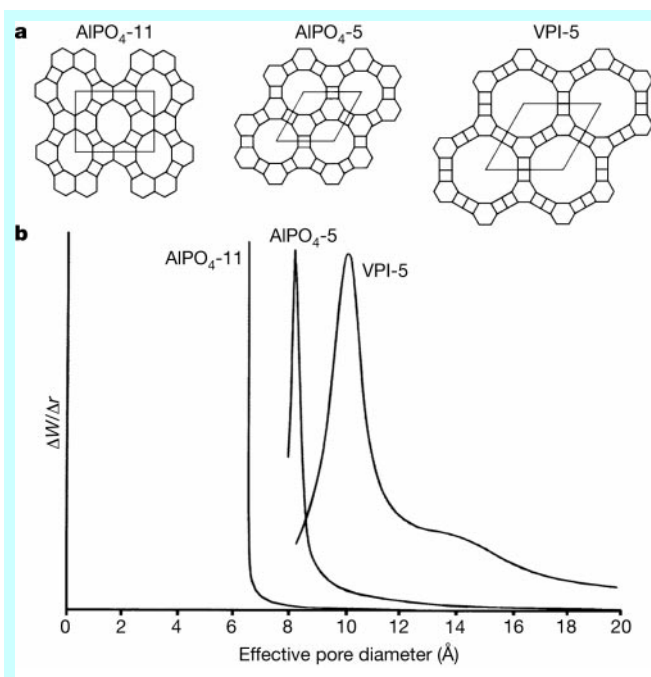


Figure 1 Pore characteristics in the aluminophosphates $\text{AlPO}_4\text{-11}$, $\text{AlPO}_4\text{-5}$ and VPI-5. **a**, Representations of the pores in the aluminophosphates $\text{AlPO}_4\text{-11}$, $\text{AlPO}_4\text{-5}$ and VPI-5. The line segments represent oxygen atoms that bridge between two tetrahedral atoms (intersection points) that are in this case either Al^{3+} or P^{5+} with strict alternation to give a composition of AlPO_4 . Rhomboids indicate the unit cell. **b**, Measured pore sizes by argon adsorption techniques. Note that VPI-5 shows a pore diameter above 1.0 nm. (ΔW is the change in adsorption mass and Δr is the change in radius of the pore size.)

uses of ever more sophisticated versions of these materials look very promising.

Microporous materials with large pores

The preparation of the aluminophosphate VPI-5—the above-mentioned crystalline microporous material with uniform pores larger than 1.0 nm—opened up the area of extra-large pore crystalline materials. Extra-large pores are obtained if more than 12 oxygen atoms span the circumference of the pore (see also Fig. 1), and the resultant pore size allows practical applications for which materials with smaller pores are not suitable. In VPI-5, the pores are circular one-dimensional channels that, owing to the crystallinity of the material, have an absolutely uniform diameter of 1.2 nm and endow the material with an approximately 30% void fraction⁸. Following the discovery of VPI-5, numerous extra-large pore materials were synthesized. Table 1 lists typical (though not all) representatives of such materials, all having structures that contain rings made of more than 12 oxygen atoms and most of them phosphate-based. Except for the silicas, these extra-large pore materials exhibit in their as-synthesized form at least one of the following features: (1) mixed metal-ion coordination (such as aluminium in octahedral and tetrahedral coordination), (2) terminal OH groups, and (3) the presence of other non-tetrahedral framework species (such as OH, H₂O, F). These features lower the framework stability relative to that of fully tetrahedrally coordinated materials such as zeolites and their pure-silica analogues. For example, JDF-20 (see Table 1) decomposes upon removal of occluded organic species⁹, and exchange of the 1,2-diaminocyclohexane molecules from as-synthesized ND-1 by alkali-metal cations and small alkylammonium cations transforms the material to other products¹⁰.

The diversity in porosity encountered in extra-large pore crystalline materials is strikingly illustrated by the phosphate-based materials listed in Table 1. For example, the pore shape of cloverite resembles a cloverleaf (hence the origin of the name)¹¹: the 20-membered ring enclosing the pore contains four terminal OH groups that intrude into the opening to provide the cloverleaf shape. Cloverite is thus unlikely to accommodate molecules of a certain size and shape that can be accommodated in the circular pores of VPI-5. In fact, the role of pore size as well as pore shape in effecting the separation of molecules has recently been demonstrated with ETS-4 (ref. 12): as illustrated schematically in Fig. 2, the pore shape of this material is crucial to discriminate between unsymmetrical N₂ and symmetrical CH₄. This principle, which has led to separation devices that are now commercially available (as Molecular Gate Technology; details available from molecular.gate@engelhard.com), might be extendable to materials with larger pores and the separation of larger molecules.

In addition to pore shape variations, the materials listed in Table 1 also exhibit pronounced differences in the shape and size of their void space. For example, the pore system of cloverite is three-dimensional, containing 3-nm cages that are each accessible through six clover-shaped pores; in contrast, VPI-5 has one-dimensional and uniform pore channels. The fact that cloverite has large

cages to which smaller pores provide access allows for the construction of inclusion species and 'ship-in-a-bottle' syntheses, where individual reagents are sufficiently small to enter through the pores, but the assembled inclusion complex is then too large to escape the cages.

The practical value of phosphate-based extra-large pore materials is limited by their relatively poor thermal and hydrothermal stabilities as compared to those of silica-based molecular sieves. Although some of the phosphate materials are sufficiently stable for certain (low-temperature) application areas, concern was raised over whether all extra-large pore materials would lack stability, owing to the presence of extra-large rings in their structures. In the case of VPI-5, the lack of stability has been attributed to the nature of the structural units, rather than to the presence of the extra-large ring (ref. 13 and references therein). This was inferred from VPI-5 and the aluminophosphate AlPO₄-H2 being made of the same basic structural units and exhibiting similar instabilities, even though the latter has one-dimensional pores comprising only 10-membered rings. More direct evidence for this interpretation was provided by the successful synthesis of the two extra-large pore silicas UTD-1 (ref. 14) and CIT-5 (ref. 15), which are both crystalline silicas containing 14-membered pore rings (see Table 1). The thermal and hydrothermal stabilities of these materials are comparable to the stabilities of other zeolites containing smaller rings within their structures, thus confirming the view that the presence of extra-large rings does not itself result in destabilization. Rather, the lack of stability seen in phosphate-based materials is due to other structural features, for example, mixed metal-ion coordination, terminal OH groups and the presence of non-tetrahedral framework species such as OH, H₂O or F (as mentioned above).

Given the limited stability of phosphate-based materials, the synthesis of extra-large pore crystalline silicas is a promising advance. Both UTD-1 and CIT-5 (see Table 1) have one-dimensional pore systems, with elliptical (0.75 × 1.0 nm) and circular (0.75 nm) pore shapes, respectively. These materials provide the precedent for the ability to synthesize extra-large pore crystalline solids and apply them to a greater variety of commercially relevant applications.

Porous metal-organic frameworks

A different approach to preparing microporous solids involves the coordination of metal ions to organic 'linker' moieties, thus yielding open framework structures. In fact, these materials have a long history, and examples include transition metal cyanide compounds (early examples are Hofmann-type clathrates, Prussian-Blue type structures and Werner complexes) and the diamond-like framework bis(adiponitrilo)copper (I) nitrate¹⁶. Open frameworks comprising metal-organic units gained renewed interest in the 1990s, but the inability of these solids to maintain permanent porosity and avoid structural rearrangements upon guest removal or guest exchange (some leading to complete collapse of the framework) has been an obvious shortcoming. However, metal-organic frameworks (MOFs) that exhibit permanent porosity have now been prepared^{17–19}. The first such solid was MOF-5, which consists of

Table 1 Representative examples of crystalline materials with ring sizes above 12

Material	Year reported	Main framework composition	Ring size (oxygen atoms)	Pore size (nm)*	Reference
VPI-5	1988	AlPO ₄	18	1.2	3, 8
AlPO ₄ -8	1990	AlPO ₄	14	<1.0	131, 132
Cloverite	1991	GaPO ₄	20	<1.0	11
JDF-20	1992	AlPO ₄	20	—†	9
ULM-5	1994	GaPO ₄	16	ND	133
UTD-1	1996	SiO ₂	14	~1.0	14
ULM-16	1996	GaPO ₄	16	ND	134
CIT-5	1997	SiO ₂	14	0.8	15
ND-1	1999	ZnPO ₄	24	ND	10
FDU-4	2001	Ge _x O _y	24	ND	135
NTHU-1	2001	GaPO ₄	24	ND	136

ND, Not determined. *By adsorption. †Structural collapse upon removal of organic.

Zn^{2+} and 1,4-benzenedicarboxylate and has a microporous volume larger than any known zeolite¹⁷. Another example is a copper-containing framework with pores of 1.64 nm diameter¹⁸. Recently developed MOF-type solids exhibit the lowest densities reported for any crystalline material ($0.41\text{--}0.21\text{ g cm}^{-3}$) as well as a high methane storage capacity¹⁹.

Porous metal–organic frameworks are unlikely to compete with zeolites and other oxide-based porous materials in high-temperature applications owing to their limited long-term stability under such conditions and high cost. But the ability to prepare such solids, including frameworks with extra-large pores and high pore volumes, is nevertheless likely to open up many application possibilities in niche areas. For example, the methane adsorption capacity of solids based on copper dicarboxylates and triethylenediamine²⁰ and of MOF-type solids¹⁹, exceeds that of any other known crystalline material. The functionalization of the organic component of the framework, or incorporation of functional organic groups directly into the framework, may yield porous solids that contain different groups capable of binding guests and/or catalysing chemical reactions involving adsorbed guests. However, organic-functionalized porous oxides are now also available²¹ and are likely to compete successfully with metal–organic frameworks in such applications.

Unique application possibilities may arise from the ability to exploit the metal component and/or its interaction with guest molecules to design porous materials with unusual physicochemical properties, such as redox potentials, light absorption properties or magnetic moments. Moreover, the metal–organic framework approach seems particularly suited to constructing a wide range of chiral porous materials^{22,23} for applications requiring enantioselective adsorption and catalysis. In contrast, the formation of chiral porous structures from oxides is proving very difficult, despite one example of enantioselective catalysis and separation using a partially chiral zeolite¹.

Ordered mesoporous materials

In parallel to the above work on crystalline microporous materials with extra-large pores and on crystalline metal–organic porous solids, the discovery and development of well-ordered, mesoporous materials occurred. Intense focus on ordered mesoporous materials

was largely initiated by a report⁴ from workers at Mobil in 1992, who described the successful preparation of mesoporous silicas with hexagonal and cubic symmetry and pore sizes ranging from 2 to 10 nm, through the use of surfactants as organizing agents. It was clear that extra-large pore, crystalline materials like VPI-5, $\text{AlPO}_4\text{-8}$ and cloverite influenced this work, and at that time, no conclusions were drawn as to whether or not the walls of the ordered mesoporous materials had precise atomic ordering⁴. In fact, the hexagonal variant of the material was speculated to be a larger-pore version of VPI-5 (see Fig. 3).

The ^{29}Si nuclear magnetic resonance (NMR) spectrum of Mobil's mesoporous material MCM-41 suggested a disordered wall structure. But this evidence is not conclusive, given that crystalline pure-silica materials with *BEA topology exhibit similar ^{29}Si NMR spectra, owing to a large number of framework defect sites. Proof for the lack of crystallinity of MCM-41 was the subsequent observation²⁴ that dehydrated samples gave Raman bands indicative of planar 3-membered ring stretch vibrations; crystalline materials containing 3-membered rings do not reveal these vibrations²⁵ because the rings are no longer isolated when contained within a continuous three-dimensional framework²⁶. Moreover, the spectra matched those of amorphous oxides and glasses with 3-membered rings at their surfaces. The lack of precise atomic positioning in these well-ordered mesoporous materials thus suggests that analogous,

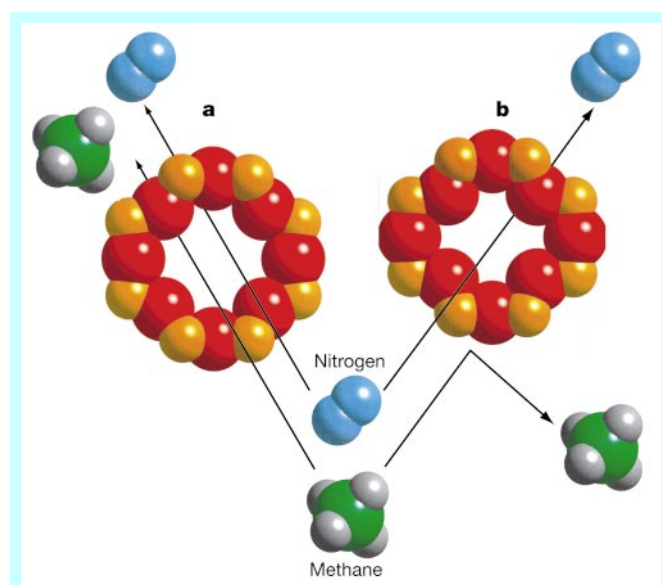


Figure 2 Pore size and shape of ETS-10. **a**, Circular pore shape does not discriminate between nitrogen and methane. **b**, Elliptical shape of pore, obtained upon heat treatment, allows nitrogen adsorption only.

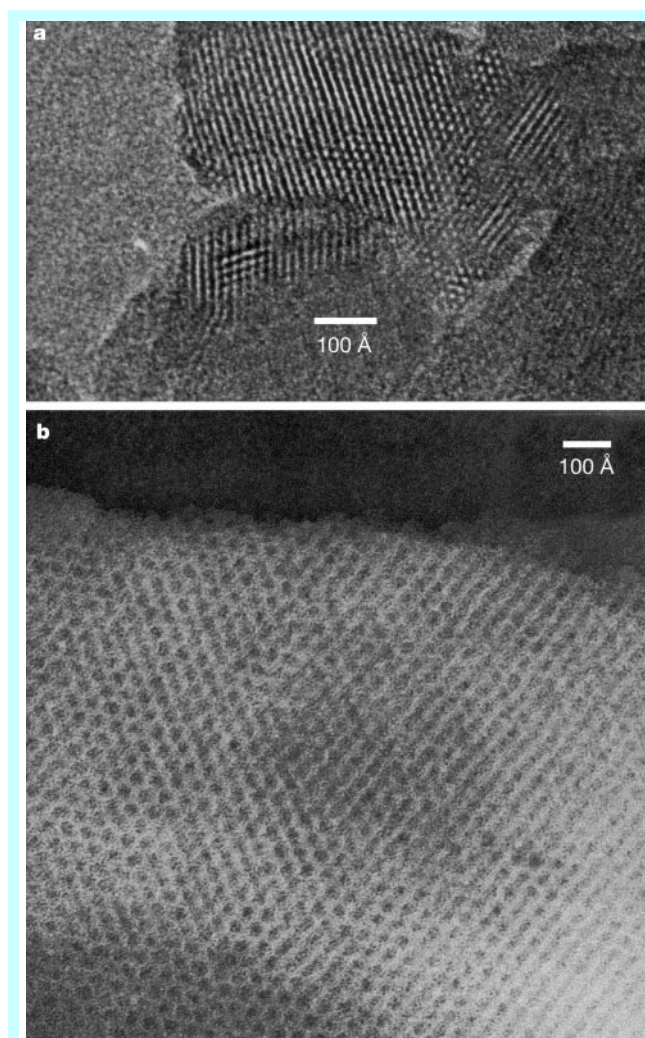


Figure 3 Transmission electron micrographs of VPI-5 and MCM-41. **a**, Micrograph of VPI-5; **b**, micrograph of MCM-41; note the difference in scale. Panel **b** is adapted from ref. 24, with permission from Elsevier Science.

well-ordered mesoporous materials can be prepared using element combinations found in other amorphous materials, such as aerogels, xerogels, organic–inorganic hybrid materials, metals and mixed-metal alloys. Indeed, some of these materials are now available in ordered, mesoporous forms^{27–29}. Very recently, an ordered, hexagonal mesoporous benzene–silica composite was prepared³⁰ from the assembly of 1-4-bis(triethoxysilyl)benzene $[(C_2H_5O)_3Si-C_6H_4-Si(OC_2H_5)_3]$ and alkytrimethylammonium surfactants. This material exhibits periodicity in the hexagonal arrangement of the mesopores obtained upon surfactant removal (5.25 nm lattice constant) and periodicity in the walls along the channel direction with a spacing of 0.76 nm (silica and aromatic units alternate). It will be interesting to see whether this unique arrangement can be generalized to hybrid materials made of other combinations of organic and inorganic units.

Although the report on MCM-41 by Kresge *et al.*⁴ stimulated recent work on ordered materials with uniform mesoporosity, evidence that such materials could be prepared precedes their publication^{31,32}. In fact, materials of the type reported by Kresge *et al.*⁴ may have already been made decades ago: in 1971, a material described as low-bulk density silica was obtained through hydrolysing and condensing tetraethylorthosilicate in the presence of cationic surfactants³³. The porosity and structure of the low-bulk density silica were not reported, but attempts to reproduce the synthesis revealed that it may have yielded ordered mesoporous materials³⁴. The first clear demonstration of the successful preparation of a material exhibiting uniform mesoporosity was achieved by Manton and Davidtz³², who obtained amorphous aluminosilicates with fairly uniform pores, by using quaternary ammonium cations such as tetrabutylammonium and tetrapropylammonium cations during the synthesis. Later, Yanagisawa *et al.* used cationic surfactants to synthesize silicates with uniform mesoporosity³¹.

The work by Kresge *et al.*⁴ was thus not in the absence of earlier claims of having prepared materials with uniform mesoporosity. But it nevertheless constituted a breakthrough because the high degree of order achieved in MCM-41 was unprecedented in the literature; there was little evidence of structural order in the materials reported before. (In fact, controlled-pore glass can have a very uniform mesopore size yet have no structural order in the glass.) Moreover, Kresge *et al.*⁴ identified the connection between the ordering observed in the mesoporous materials and the structure-directing aggregation properties of the surfactants used in their synthesis. (Block copolymers also form mesoscopic assemblies and are now used in the preparation of ordered mesoporous materials.) The combination of these two contributions—the novelty of the new materials, with their unprecedented high degree of ordering, and the introduction of a clear conceptual framework that could guide the design of new materials—stimulated an enormous amount of work. Today, a large number of mesoporous materials of varying composition, pore size and inorganic wall thickness are available.

With the exception of the ordered, hexagonal mesoporous benzene–silica composite prepared by Inagaki *et al.*³⁰, ordered mesoporous materials are not crystalline, thus permitting their

synthesis by many different routes, as illustrated by the examples reported in references 35–39. As might be expected, the degree of ordering can depend significantly on the method used to synthesize the mesoporous material, ranging from the high degree of ordering first reported by Kresge *et al.*⁴ to an almost complete lack of ordering. That is, while uniformly sized mesopores are produced, the degree to which they form ordered arrangements within the material shows tremendous variability.

It remains unclear why the vast majority of ordered mesoporous materials cannot be synthesized in crystalline form. The recent example of structural order in the walls of the mesoporous benzene–silica composite³⁰ may help shed light on this issue. Regarding the inorganic mesoporous materials, their syntheses occur under thermodynamic conditions comparable to those encountered during the syntheses of the more open zeolite frameworks⁴⁰, suggesting that thermodynamic barriers in the assembly process are not causing the lack of crystallinity. Ordered mesoporous materials were first prepared using charged organic components, and charge matching between the organic and inorganic components should influence the assembly of ordered mesoporous solids. This aspect of their synthesis may be an impediment to crystallinity⁴¹. However, the ordered mesoporous materials can now be prepared using non-ionic organic components. These assemblies would suggest that charge matching is not the only impediment to crystallinity. Likewise, it is now known that not only ordered mesoporous materials can form through intermediate phases that are layered, but also zeolites (examples include MCM-22 (ref. 42) and ferrierite⁴³). The involvement of layered intermediates during formation thus does not seem to be why ordered mesoporous materials do not crystallize.

A plausible explanation for the lack of crystallinity may lie in the correlation between the framework density (FD, the number of atoms per nm³) and the structural features of porous materials. In 1989, Brunner and Meier revealed that crystalline structures containing tetrahedral atoms (T-atoms) exhibit a correlation between their FD and the parameter MINR, the minimum ring size for every T-atom⁴⁴. (If the structure has T-atoms in rings of different sizes, then a “+” is assigned. For example, MINR = 4+ denotes that some of the T-atoms are in 4-membered rings, while others are in rings of larger size.) All crystalline and ‘tetrahedral’ materials follow this relationship. In the case of oxide-based solids, most structures have MINR ≥ 4, which would correlate with a maximum void fraction of about 0.5. The correlation does not hold for crystalline structures built from units containing non-tetrahedral atoms, which have indeed been shown to exceed this void fraction^{17–19}. Regardless of the phenomena that are the origin of the correlation between the FD and MINR, they may be the reasons why the ordered mesoporous materials are not crystalline (void fractions violate this relationship).

In order to obey the correlation, ordered mesoporous materials would need to have MINR = 3+; that is, they would need to contain a large number of three-membered rings (3MRs). The recognition that 3MR building units are essential if very open framework materials are to be realized⁴⁵ initially focused attention on beryllosilicate materials because many beryllosilicate minerals, albeit dense, have a large number of 3MR. In fact, the successful synthesis⁴⁶ of an analogue of the mineral lovdarite illustrated that synthetic materials with a high number of 3MRs could be prepared. However, beryllium is highly toxic, and practically more useful implementations of the ‘3MR concept’ based on the FD-MINR relationship will therefore depend on the development of synthetic routes to 3MR structures using elements other than beryllium⁴⁷.

Zinc seems a promising and non-toxic substitute for beryllium, given that some dense zincosilicates have isostructures to beryllosilicates that contain a large number density of 3MRs. Zincosilicate syntheses aimed specifically at the preparation of microporous crystalline solids containing 3MRs have therefore been developed^{48,49}.

Table 2 Synthetic framework materials with numerous three-membered rings

Material	Composition	FD (atoms per nm ³)	Pore size	Year	Reference
Lovdarite	Beryllosilicate	18.4	9 MR	1986	46
VPI-7	Zincosilicate	17.1	9 MR	1991	48, 49
RUB-17	Zincosilicate	16.8	9 MR	1995	137
VPI-9	Zincosilicate	16.7	8 MR	1996	138
VPI-10	Zincosilicate	15.3	9 MR	1996	139
RUB-23	Lithosilicate	17.7	8 MR	2000	140
RUB-29	Lithosilicate	17.7	10 MR	2001	141
ASU-15	Ge ₂ ZrO ₆ F	9.1	10 MR	2000	142
OSB-1	Beryllosilicate	13.3	14 MR	2001	50
OSB-2	Beryllosilicate	12.7	8 MR	2001	50

FD, framework density; MR, membered rings.

Table 2 lists the synthetic crystalline materials that possess a significant number of 3MRs, which include zincosilicates, lithium silicates and a $\text{Ge}_2\text{ZrO}_6\text{F}$ material. The materials listed illustrate the significant progress in the synthesis of MINR = 3+ materials, which now provides routes to new porous solids. Recently, an extra-large pore beryllsilicate with MINR = 3 has been reported⁵⁰, which constitutes a promising extension of the '3MR concept' of Meier⁴⁵. Although beryllium's toxicity precludes practical use of this material, the conversion to a zincosilicate isostructure should be feasible.

As stated above, the reason that ordered mesoporous materials cannot be synthesized in a crystalline form is unknown. Materials with the pore volumes of the ordered mesoporous materials would require MINR = 3+ if they were to be tetrahedral frameworks, so it is possible that the appropriate oxide chemistry that would allow the realization of such materials has not yet been explored. A recently reported ordered mesoporous zincosilicate⁵¹ did not contain sufficient amounts of zinc to facilitate the formation of a significant number of 3MRs, and further exploration of zinc-based materials might thus allow the preparation of ordered mesoporous materials that are crystalline.

Hierarchical structures

Interest in hierarchical porous structures has burgeoned over the past decade. Zeolite particles have been used as building blocks to construct hierarchical porous structures. Pure, crystalline molecular sieves have been produced as free-standing films⁵², and films, spheres and fibres have been constructed from nanometre-sized crystals^{53–55}. Techniques that use colloidal suspensions to shape ceramic structures are well established, and their extension to microporous colloidal particles for casting gels into designed shapes⁵⁶ and for film casting⁵⁷ provides flexible synthetic routes to a range of hierarchical microporous structures.

In addition to the use of such 'forming operations', hierarchical structures have also been created through bulk dissolution and structural rearrangement approaches. For example, tubes and fibres of zeolite have been prepared by conversion of pre-existing oxide tubes and fibres⁵⁸. An interesting illustration is the zeolitization of diatoms, which involved conversion of a portion of the diatom silica structure into the zeolite ZSM-5, to yield a micro- and mesoporous hierarchical structure⁵⁹. (Diatoms are single-celled algae enclosed by silica shells that, upon death of the algae, collect on the ocean/lake bed as diatomaceous earth or kieselguhr, a material widely used in filtration applications.).

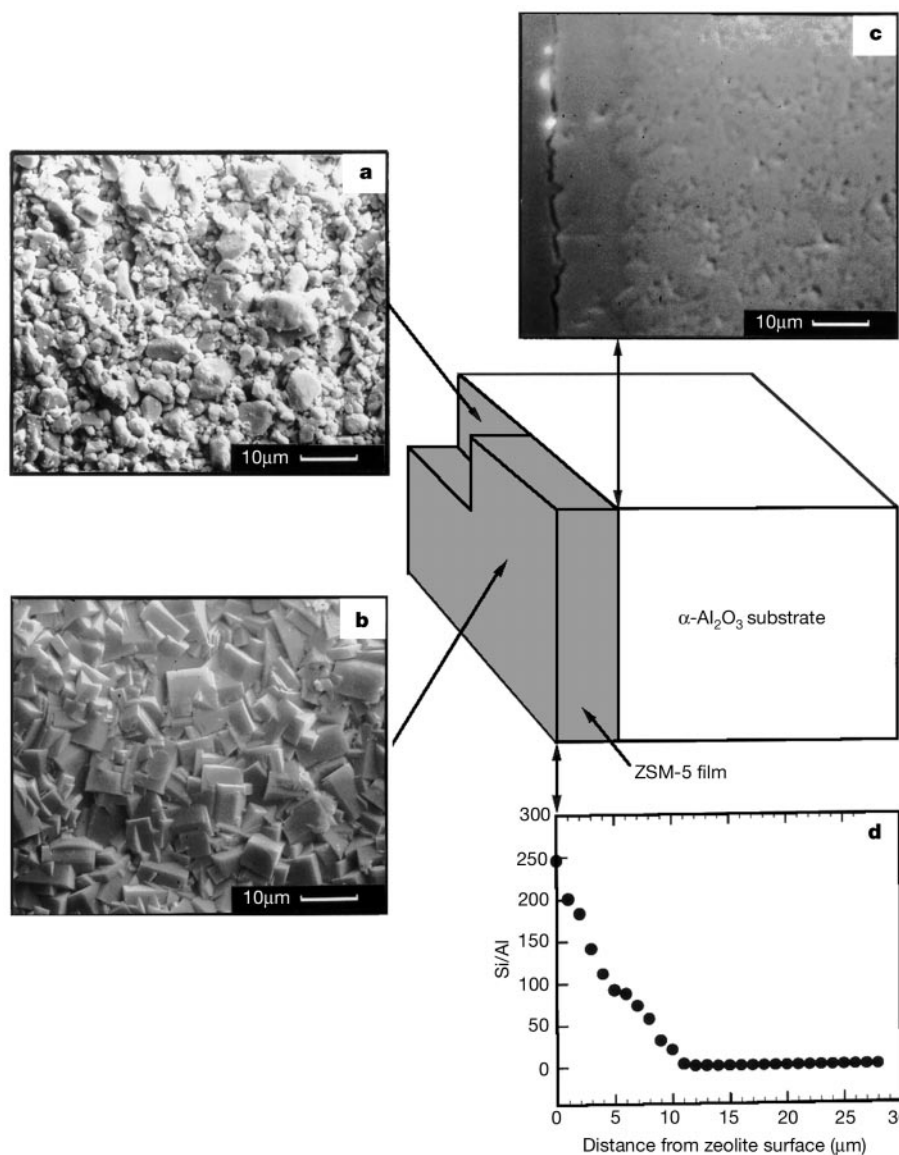


Figure 4 Characteristics of a supported zeolite ZSM-5 film. Shown are scanning electron micrographs of different surfaces (**a,b,c**) of polycrystalline zeolite ZSM-5 film supported on a porous $\alpha\text{-Al}_2\text{O}_3$ substrate and (**d**) results of electron microprobe analysis showing the zeolite layer to be approximately 10 μm in thickness (consistent with image shown in **c**, which is a view from the top). Adapted from ref. 80; reproduced by permission of The Royal Society of Chemistry.

Hierarchical structures can also be prepared from mesoporous materials. Films, fibres and spheres possessing fairly uniform mesoporosity have been fabricated using surfactants as structure-directing agents, in a manner similar to the synthesis of ordered mesoporous materials^{60,61}.

Hierarchical porous solids possessing ordered pores on the micro-, and meso- and even macroscale^{62,63} have been developed. The hierarchical structure of these materials is expected to open new application opportunities. Until now, most progress has been achieved with films, in terms of both our scientific understanding of their formation and the ability to exploit them for practical applications.

The first zeolite-based films, both supported⁶ and freestanding⁵², were reported by the early 1990s. Zeolite thin films (Fig. 4) and layers (films and layers refer to continuous and discontinuous structures, respectively; ref. 64 and references therein) are of interest for use as membranes in separations devices and membrane reactors. They can also serve as chemical sensors and hosts for guest species, which in turn impart the material with optical, electrical or magnetic properties.

Early work on zeolite and molecular sieve layers^{65–68}, including their use as chemical sensors^{65,68}, yielded layers of randomly oriented materials, while more recent examples contain oriented metallophosphates (zincophosphate, $\text{AlPO}_4\text{-5}$) and zeolite A. The oriented $\text{AlPO}_4\text{-5}$ crystals were prepared on alumina membranes⁶⁹, while oriented zeolites were obtained using microstructured silicon wafers⁷⁰ and metal films⁷¹. Recently, the heteroepitaxial growth of the zeolites cancrinite^{72,73} and chabazite⁷³ from the surface of millimetre-sized sodalite crystals has been achieved. This breakthrough in zeolite synthesis opens the way to constructing microporous assemblies with precise pore orientation over sufficiently large length scales to be useful for device applications, such as in optics. Oriented layers may find potential use as sensors and other devices, but the lack of continuity precludes their use as membranes. Continuous films have been prepared for use in sensors⁷⁴ and as low dielectric constant materials⁷⁵, but whether a continuous thin film is indeed required will depend on the particular application^{64–68}.

In the 1980s, attempts to prepare continuous zeolite films showed little success, but the past decade has seen numerous demonstrations that molecular sieving membranes can be fabricated on ceramic and metal supports in disk and tubular configurations^{76–81}. These films usually contained unaligned ZSM-5 zeolite, but continuous and oriented zeolite films have now also been prepared^{82–85} and the orientation shown to affect the permeation and molecular sieving behaviour of the membranes⁸⁶. Membranes synthesized

from other zeolites (such as zeolite A, zeolite Y, mordenite, ferrierite and zeolite UTD-1) and molecular sieves (such as $\text{AlPO}_4\text{-5}$) are now available as well.

In addition to molecular sieve films and layers, ordered mesoporous materials have been fabricated in these configurations^{87–96}. The mesoporous films have been investigated as hosts for materials that act as pH sensors^{90,91}, shown to function as low dielectric barriers^{87–89,92–94} and incorporated as functional elements in photonic devices⁹⁴. For some mesoporous materials, orientation of the mesopore system with respect to the supporting substrate has been achieved^{95,96}.

New applications

Over the past decade crystalline microporous materials have continued to find new applications in their traditional areas of use, such as catalysis, separation and ion exchange. Following the discovery of ordered mesoporous materials, these solids have also been regarded as suitable for such applications, with several studies illustrating that they can serve as stationary phases in high performance liquid chromatography^{97–99}. Although biomolecule separations have not yet been attempted, the pore size and pore uniformity of mesoporous materials may make them suitable for macromolecule separations, which, if accomplished, should lead to commercial use.

In addition, non-traditional applications exploiting the porosity of the material itself, or its ability to host guest species that impart an interesting property, are now also nearing commercialization, or show high potential to do so. For example, microchip device densities continue to increase and feature sizes to decrease, generating much demand for insulators with a lower dielectric constant k than that of dense silicon dioxide ($k = 3.9\text{--}4.2$), the currently used insulator material¹⁰⁰. Porous materials can provide for low- k materials ($k < 2.2$)¹⁰¹, as demonstrated by the performance of films prepared from ordered mesoporous materials^{87–89,92,93} and crystalline pure-silica molecular sieves^{75,102}. In fact, one mesoporous silica material, MesoELK, is reported to be targeted for commercial use in 2002, in applications requiring low dielectric constant films (see <http://www.schumacher.com/products/DESCRIP/CHEM/Meso.htm> for details).

Another promising application area is diagnostic magnetic resonance imaging (MRI), which relies on administering contrast agents to patients to significantly improve the diagnostic value of MRI. The contrast agents contain high-spin metals that bind water molecules and thereby yield proton spin relaxation times (T_1) that are orders of magnitude faster than those obtained with free water¹⁰³. Gadolinium ions, Gd^{3+} , perform particularly well as contrast agents, but cannot be administered directly owing to their inherent toxicity¹⁰⁴. However, Gd^{3+} contained in zeolites^{105,106} and clays^{107,108} has been investigated as contrast agent for the gastrointestinal tract. In fact, a zeolite containing Gd^{3+} is nearing completion of the regulatory process and is expected to be commercially available soon, marketed as Citra Vu by EZM, Canada (K. J. Balkus Jr, personal communication). The zeolite immobilizes the Gd^{3+} and thus mitigates its toxicity (zeolites themselves are not toxic when introduced in the gastrointestinal tract), while still allowing hydration by adsorbed water to allow for NMR T_1 relaxation effects.

Another interesting example among the many uses of porous solids as a host material is the zeolite-dye microlaser: incorporation of organic dyes as guest species in microporous, crystalline $\text{AlPO}_4\text{-5}$ (Fig. 5) gave the first lasing molecular sieve^{109,110}, with size, shape and arrangement of the porous crystals affecting the performance of the system^{109,111}. The concept has been extended to mesoporous materials acting as the host for laser dyes^{112,113}, and the use of other guest species as light-emitters, such as complexed Nd^{3+} , is also explored¹¹⁴.

A somewhat different approach to using porous solids as host material involves the fabrication of porous carbons, which are of

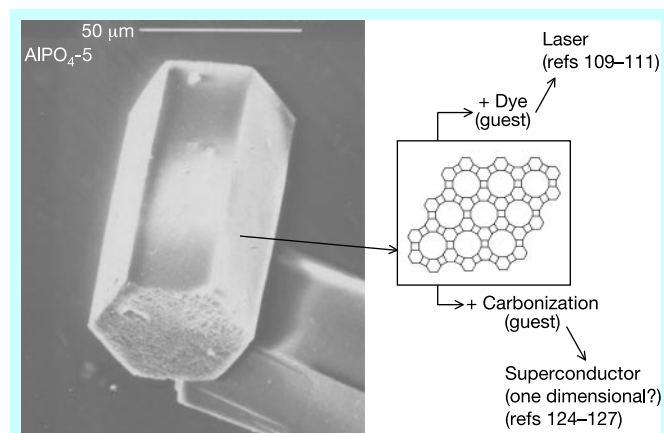


Figure 5 The use of $\text{AlPO}_4\text{-5}$ as a host for guest molecules. Shown are a micrograph of a $\text{AlPO}_4\text{-5}$ microcrystal and a schematic representation of its pore structure that can serve as a host for the inclusion of different guest species, to endow the composite structures with new properties.

interest as adsorbents, electrode material, and so on. It involves incorporation of organic compounds such as acrylonitrile¹¹⁵, poly-(acrylonitrile)¹¹⁶, poly(furfuryl alcohol)^{116,117} or phenolformaldehyde polymers¹¹⁸ in zeolites such as Y^{115–118}, mordenite¹¹⁵, beta¹¹⁸ or L¹¹⁸, followed by carbonization and subsequent removal of the inorganic material by acid dissolution. The porous carbons obtained in this manner usually lack uniform pore structure, but one material appears to have structural order with highly uniform porosity¹¹⁷. This fabrication method has been extended to using ordered mesoporous materials as hosts^{119–123}, yielding rigid arrays of nanoporous carbon with pore sizes ranging from 5 to 10 nm. Like the ordered mesoporous oxide-based materials, the mesoporous carbons obtained in this manner can be highly ordered¹²³. These carbon-based porous solids are likely to exhibit material properties not encountered in other mesoporous solids.

The pores of a crystalline molecular sieve have also been used as a confining reaction environment to produce single-walled carbon nanotubes (Fig. 5), through the carbonization of occluded tripropylamine^{124,125}. These tubes are claimed to be merely 0.4 nm in diameter^{124,126} and to show superconductivity¹²⁵ that may be one-dimensional¹²⁷. Further work is necessary to understand the precise nature of this composite material, and to establish whether carbon nanotubes can be formed in other porous structures.

Future challenges

As mentioned above, it is unknown why ordered mesoporous materials do not form crystalline frameworks. Crystalline structures containing pores in the size range of 1.0–2.0 nm and exhibiting thermal and hydrothermal stability are highly desirable, especially if the pore system is more than one-dimensional. A major challenge facing the porous materials community is therefore the development of tetrahedral frameworks containing extra-large pores in a multidimensional network. The largely unexplored approach of creating open structures that are based on 3-membered rings might prove successful in this regard. In addition, new types of organic molecules that can act as structure-directing agents are being developed, not least because these costly reagents can often limit the commercial viability of the resultant porous solid. A promising development is the preparation of such organic molecules that can be recycled¹²⁸.

The synthesis of an enantiopure, porous tetrahedral framework constitutes another challenge that is not new, but remains unsolved. Among the criteria for synthesizing such a chiral solid¹ is the need for chiral organic molecules as templates. But suitable compounds have not yet been found: the structures obtained when using occluded chiral organic agents exhibit disorder in the organic. Recently, the ordering of fluoride counter ions in a molecular sieve was found to affect the ordering of the organic residing in the porous space of the sieve¹²⁹, giving rise to a fluoride-ion–organic arrangement with noncentrosymmetric topological symmetry and thus nonlinear optical behaviour. This effect, in combination with the use of a chiral template, might provide a route to chiral molecular sieves.

Zeolites and other micro- and mesoporous materials are certain to continue to find new uses in catalysis¹³⁰ and separation applications¹². Metal–organic frameworks might also find use in such applications, particularly as they are likely to offer advantages over other porous solids for catalysing chiral transformations that do not require high temperatures. The preparation and characterization of chiral metal–organic frameworks seems therefore particularly promising, especially when aimed at creating pore sizes above 1.0 nm (many useful reactants for chiral transformations are nanometre-sized or larger).

As better and more diverse preparations of hierarchical porous structures become available, an increasing number of porous materials are likely to find use in applications not traditionally associated with such materials. The recent demonstration of hetero-

epitaxial zeolite growth is only one promising example of developments that could lead to new device applications.

Clearly, numerous challenges remain, but the rate of advance in porous materials science over the past decade promises that this area will deliver exciting developments in the early twenty-first century. □

doi:10.1038/nature00785.

- Davis, M. E. & Lobo, R. F. Zeolite and molecular sieve synthesis. *Chem. Mater.* **4**, 756–768 (1992).
- Corma, A. From microporous to mesoporous molecular sieve materials and their use in catalysis. *Chem. Rev.* **97**, 2373–2419 (1997).
- Davis, M. E., Saldarriaga, C., Montes, C., Garces, J. & Crowder, C. A molecular sieve with eighteen-membered rings. *Nature* **331**, 698–699 (1988).
- Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism. *Nature* **359**, 710–712 (1992).
- Hoskins, B. F. & Robson, R. Design and construction of a new class of scaffolding-like materials comprising infinite polymeric frameworks of 3-D-linked molecular rods—A reappraisal of the Zn(CN)₂ and Cd(CN)₂ structures and the synthesis and structure of the diamond-related frameworks [N(CH₃)₄][CuZnII(CN)₄] and CuI[4,4',4'',4'''-tetracyanotetra-phenylmethane]BF₄·XC₆H₅NO₂. *J. Am. Chem. Soc.* **112**, 1546–1554 (1990).
- Suzuki, H. Composite membrane having a surface layer of an ultrathin film of cage-shaped zeolite and process for production thereof. US Patent 4,699,892 (1987).
- Stucky, G. D. & MacDougall, J. E. Quantum confinement and host guest chemistry—Probing a new dimension. *Science* **247**, 669–678 (1990).
- Davis, M. E. *et al.* Physicochemical properties of VPI-5. *J. Am. Chem. Soc.* **111**, 3919–3924 (1989).
- Huo, Q. H. *et al.* Synthesis and characterization of a novel extra large ring of aluminophosphate JDF-20. *J. Chem. Soc. Chem. Commun.* 875–876 (1992).
- Yang, G. Y. & Sevon, S. C. Zinc phosphate with gigantic pores of 24 tetrahedra. *J. Am. Chem. Soc.* **121**, 8389–8390 (1999).
- Estermann, M., McCusker, L. B., Baerlocher, Ch., Merrouche, A. & Kessler, H. A synthetic gallophosphate molecular-sieve with a 20-tetrahedral-atom pore opening. *Nature* **352**, 320–323 (1991).
- Kuznicki, S. M. *et al.* A titanosilicate molecular sieve with tunable pores and its use in gas separation. *Nature* **412**, 720–724 (2001).
- Li, H. X. & Davis, M. E. Phosphate-based molecular sieves with pores comprised of greater than 12-rings. *Catal. Today* **19**, 61–106 (1994).
- Lobo, R. F. *et al.* Characterization of the extra-large-pore zeolite UTD-1. *J. Am. Chem. Soc.* **119**, 8474–8484 (1997).
- Yoshikawa, M. *et al.* Synthesis, characterization and structure solution of CIT-5, a new, high-silica, extra-large-pore molecular sieve. *J. Phys. Chem. B* **102**, 7139–7147 (1998).
- Kinoshita, Y., Matsubara, I., Higuchi, T. & Saito, Y. The crystal structure of bis(adiponitrile)copper (I) nitrate. *Bull. Chem. Soc. Jpn* **32**, 1221–1226 (1959).
- Li, H., Eddaoudi, M., O'Keeffe, M. & Yaghi, O. M. Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature* **402**, 276–279 (1999).
- Chen, B., Eddaoudi, M., Hyde, S. T., O'Keeffe, M. & Yaghi, O. M. Interwoven metal-organic framework on a periodic minimal surface with extra-large pores. *Science* **291**, 1021–1023 (2001).
- Eddaoudi, M. *et al.* Systematic design of pore size and functionality in isorecticular MOFs and their application in methane storage. *Science* **295**, 469–472 (2002).
- Seki, K. Design of an adsorbent with an ideal pore structure for methane adsorption using metal complexes. *Chem. Commun.* 1496–1497 (2001).
- Jones, C. W., Tsuji, K. & Davis, M. E. Organic-functionalized molecular sieves as shape-selective catalysts. *Nature* **393**, 52–54 (1998).
- Seo, J. S. *et al.* A homochiral metal-organic porous material for enantioselective separation and catalysis. *Nature* **404**, 982–986 (2000).
- Keper, C. J., Prior, T. J. & Rosseinsky, M. J. A versatile family of interconvertible microporous chiral molecular frameworks: The first example of ligand control of network chirality. *J. Am. Chem. Soc.* **122**, 5158–5168 (2000).
- Chen, C. Y., Li, H. X. & Davis, M. E. Studies on mesoporous materials. I. Synthesis and characterization of MCM-41. *Microporous Mater.* **2**, 17–26 (1993).
- Annen, M. J. & Davis, M. E. Raman and ²⁹Si MAS NMR spectroscopy of framework materials containing three-membered rings. *Microporous Mater.* **1**, 57–65 (1993).
- De Man, A. J. M., Ueda, S., Annen, M. J., Davis, M. E. & van Santen, R. A. The stability and vibrational spectra of three-ring containing zeolitic silica polymorphs. *Zeolites* **12**, 789–800 (1992).
- Huo, Q. *et al.* Generalized synthesis of periodic surfactant inorganic composite-materials. *Nature* **368**, 317–321 (1994).
- Inagaki, S., Guan, S., Fukushima, Y., Ohsuna, T. & Terasaki, O. Novel mesoporous materials with a uniform distribution of organic groups and inorganic oxide in their frameworks. *J. Am. Chem. Soc.* **121**, 9611–9614 (1999).
- Attard, G. S. *et al.* Mesoporous Pt/Ru alloy from the hexagonal lyotropic liquid crystalline phase of a nonionic surfactant. *Chem. Mater.* **13**, 1444–1446 (2001).
- Inagaki, S., Guan, S., Ohsuna, T. & Terasaki, O. Mesoporous organic-silica hybrid with crystal-like pore walls. *Nature* **416**, 304–307 (2002).
- Yanagisawa, T., Shimizu, T., Kuroda, K. & Kato, C. The preparation of alkyltrimethylammonium-kanemite complexes and their conversion to microporous materials. *Bull. Chem. Soc. Jpn* **63**, 988–992 (1990).
- Manton, M. R. S. & Davidtz, J. C. Controlled pore sizes and active site spacings determining selectivity in amorphous silica-alumina catalysts. *J. Catal.* **60**, 156–166 (1979).
- Chiola, V., Ritsko, J. E. & Vanderpool, C. D. US Patent 3,556,725 (1971).
- Di Renzo, F., Cambon, H. & Dutarte, R. A 28-year-old-synthesis of micelle-templated mesoporous silica. *Microporous Mater.* **10**, 283–286 (1997).
- Inagaki, S., Fukushima, Y. & Kuroda, K. Synthesis of highly ordered mesoporous materials from a layered polysilicate. *J. Chem. Soc. Chem. Commun.* 680–682 (1993).
- Chen, C. Y., Xiao, S. Q. & Davis, M. E. Studies on ordered mesoporous materials. III. Comparison of

- MCM-41 to mesoporous materials derived from kanemite. *Microporous Mater.* **4**, 1–20 (1995).
37. Sakamoto, Y. *et al.* Structure analysis of mesoporous material 'FSM-16'—Studies by electron microscopy and X-ray diffraction. *Microporous Mesoporous Mater.* **21**, 589–596 (1998).
38. Kimura, T. *et al.* Formation of novel ordered mesoporous silicas with square channels and their direct observation by transmission electron microscopy. *Angew. Chem. Int. Edn Engl.* **39**, 3855–3859 (2000).
39. Landry, C. C. *et al.* Phase transformations in mesostructured silica/surfactant composites. Mechanisms for change and applications to materials synthesis. *Chem. Mater.* **13**, 1600–1608 (2001).
40. Navrotsky, A., Petrovic, I., Hu, Y., Chen, C. Y. & Davis, M. E. Little energetic limitation to microporous and mesoporous materials. *Microporous Mater.* **4**, 95–98 (1995).
41. Davis, M. E., Chen, C. Y., Burkett, S. L. & Lobo, R. L. Synthesis of (aluminosilicate) materials using organic molecules and self-assembled organic aggregates as structure-directing agents. *Mater. Res. Soc. Symp. Ser. Proc.* **346**, 831–842 (1994).
42. Leonowicz, M. E., Lawton, J. A., Lawton, S. L. & Rubin, M. K. MCM-22, a molecular-sieve with 2 independent multidimensional channel systems. *Science* **264**, 1910–1913 (1994).
43. Schreyeck, L., Caullet, P., Mougénel, J. C., Guth, J. L. & Maler, B. A layered microporous aluminosilicate precursor of FER-type zeolite. *J. Chem. Soc. Chem. Commun.* 2187–2188 (1995).
44. Brunner, G. O. & Meier, W. M. Framework density distribution of zeolite-type tetrahedral nets. *Nature* **337**, 146–147 (1989).
45. Meier, W. M. Zeolites and zeolite-like materials. *Stud. Surf. Sci. Catal.* **28**, 13–22 (1986).
46. Ueda, S., Koizumi, M., Baerlocher, Ch., McCusker, L. B. & Meier, W. M. 7th Int. Zeolite Conf., Tokyo, Poster Paper 3C-3 (1986).
47. Davis, M. E. Multidimensional large pores. *Nature* **337**, 117 (1989).
48. Annen, M. J., Davis, M. E., Higgins, J. B. & Schlenker, J. L. VPI-7: The first zirconosilicate molecular sieve containing three-membered T-atom rings. *J. Chem. Soc. Chem. Commun.* 1175–1176 (1991).
49. Annen, M. J., Davis, M. E., Higgins, J. B. & Schlenker, J. L. The physicochemical properties of VPI-7: A microporous zirconosilicate with three-membered rings. *Mater. Res. Soc. Symp. Ser. Proc.* **233**, 245–253 (1991).
50. Cheetham, A. *et al.* Very open microporous materials. From concept to reality. *Stud. Surf. Sci. Catal.* **135**, [CD-ROM] Paper 05-O-05 (Elsevier, 2001).
51. Katovic, A. *et al.* Preparation and characterization of mesoporous molecular sieves containing Al, Fe or Zn. *Microporous Mesoporous Mater.* **44–45**, 275–281 (2001).
52. Haag, W. O. & Tsikoyiann, J. G. Membrane composed of a pure molecular sieve. US Patent 5,019,263 (1991).
53. Wang, X. D. *et al.* Fabrication of hollow zeolite spheres. *Chem. Commun.* 2161–2162 (2000).
54. Rhodes, K. H., Davis, S. A., Caruso, F., Zhang, B. J. & Mann, S. Hierarchical assembly of zeolite nanoparticles into ordered macroporous monoliths using core-shell building blocks. *Chem. Mater.* **12**, 2832–2834 (2000).
55. Huang, L. M. *et al.* Fabrication of ordered porous structures by self-assembly by zeolite nanocrystals. *J. Am. Chem. Soc.* **122**, 3530–3531 (2000).
56. Wang, H., Huang, L., Wang, Z., Mitra, A. & Yan, Y. Hierarchical zeolite structures with designed shape by gel-casting of colloidal nanocrystal suspensions. *Chem. Commun.* 1364–1365 (2001).
57. Tsapatsis, M., Okubo, T., Lovallo, M. & Davis, M. E. Synthesis and structure of ultrafine zeolite KL (LTL) crystallites and their use for thin film zeolite processing. *Mater. Res. Soc. Symp. Ser. Proc.* **371**, 21–26 (1995).
58. Shimizu, S. & Hamada, H. Direct conversion of bulk materials into MFI zeolites by a bulk-material dissolution technique. *Adv. Mater.* **12**, 1332–1335 (2000).
59. Anderson, M. W., Holmes, S. M., Hanif, N. & Cundy, C. S. Hierarchical pore structures through diatom zeolitization. *Angew. Chem. Int. Edn Engl.* **39**, 2707–2710 (2000).
60. Schacht, S., Huo, Q., Voigt-Martin, I. G., Stucky, G. D. & Schüth, F. Oil-water interface templating of mesoporous macroscale structures. *Science* **273**, 768–771 (1996).
61. Bruinsma, P. J., Kim, A. Y., Liu, J. & Baskaran, S. Mesoporous silica synthesized by solvent evaporation: Spun fibers and spray-dried hollow spheres. *Chem. Mater.* **9**, 2507–2512 (1997).
62. Velez, O. D., Jede, T. A., Lobo, R. F. & Lenhoff, A. M. Porous silica via colloidal crystallization. *Nature* **389**, 447–448 (1997).
63. Stein, A. Sphere templating methods for periodic porous solids. *Microporous Mesoporous Mater.* **44–45**, 227–239 (2001).
64. Bein, T. Synthesis and applications of molecular sieve layers and membranes. *Chem. Mater.* **8**, 1636–1653 (1996).
65. Yan, Y. & Bein, T. Molecular sieve sensors for selective ethanol detection. *Chem. Mater.* **4**, 975–977 (1992).
66. Feng, S. & Bein, T. Growth of oriented molecular-sieve crystals on organophosphonate films. *Nature* **368**, 834–836 (1994).
67. Feng, S. & Bein, T. Vertical aluminophosphate molecular-sieve crystals grown at inorganic-organic interfaces. *Science* **265**, 1839–1841 (1994).
68. Mintova, S., Mo, S. Y. & Bein, T. Humidity sensing with ultrathin LTA-type molecular sieve films grown on piezoelectric devices. *Chem. Mater.* **13**, 901–905 (2001).
69. Wu, C. N., Chao, K. J., Tsai, T. G., Chiou, Y. H. & Shih, H. C. Oriented growth of molecular sieves on inorganic membranes. *Adv. Mater.* **8**, 1008–1012 (1996).
70. Scandella, L., Binder, G., Gobrecht, J. & Jansen, J. C. Alignment of single-crystal zeolites by means of microstructured surfaces. *Adv. Mater.* **8**, 137–139 (1996).
71. Caro, J. *et al.* Aligned molecular-sieve crystals. *Adv. Mater.* **4**, 273–276 (1992).
72. Okubo, T. *et al.* Heteroepitaxial growth of a zeolite. *Angew. Chem. Int. Edn Engl.* **40**, 1069–1071 (2001).
73. Wakiyama, T. *et al.* Heteroepitaxial connection of zeolites with different pore structures. *Stud. Surf. Sci. Catal.* **135**, [CD-ROM] Paper 02-P-28 (Elsevier, 2001).
74. Sun, J. T., Datt, C. B. & Davis, M. E. Molecular sieve coated SAW device for the detection of carbon dioxide in the presence of water. *Mater. Res. Soc. Symp. Ser. Proc.* **360**, 359–366 (1995).
75. Wang, Z. B., Wang, H. T., Mitra, A., Huang, L. M. & Yan, Y. S. Pure-silica zeolite low- k dielectric thin films. *Adv. Mater.* **13**, 746–749 (2001).
76. Geus, E. R., van Bekkum, H., Bakker, W. J. W. & Mouljijn, J. A. High-temperature stainless steel supported zeolite (MFI) membranes: Preparation, module construction and permeation experiments. *Microporous Mater.* **1**, 131–147 (1993).
77. Jia, M. D., Peinemann, K. V. & Behling, R. D. Ceramic zeolite composite membranes. Preparation, characterization and gas permeation. *J. Membr. Sci.* **82**, 15–26 (1993).
78. Jia, M. D., Chen, B., Noble, R. D. & Falconer, J. Ceramic-zeolite composite membranes and their application for separation of vapor/gas mixtures. *J. Membr. Sci.* **90**, 1–10 (1994).
79. Matsukata, M., Nishiyama, N. & Ueyama, K. Zeolitic membrane synthesized on a porous alumina support. *J. Chem. Soc. Chem. Commun.* 339–340 (1994).
80. Yan, Y. H., Tsapatsis, M., Gavalas, G. R. & Davis, M. E. Zeolite ZSM-5 membrane grown on porous α - Al_2O_3 . *J. Chem. Soc. Chem. Commun.* 227–228 (1995).
81. Vroon, Z. A. E. P., Keizer, K., Gilde, M. J., Verweij, H. & Burggraaf, A. J. Transport properties of alkanes through ceramic thin zeolite MFI membranes. *J. Membr. Sci.* **113**, 293–300 (1996).
82. Lovallo, M. C. & Tsapatsis, M. Preferentially oriented sub-micron silicalite membranes. *AIChE J.* **42**, 3020–3029 (1996).
83. Boudreau, L. & Tsapatsis, M. A highly oriented thin film of zeolite A. *Chem. Mater.* **9**, 1705–1709 (1997).
84. Balkus, K. J. Jr, Muñoz, T. & Gimón-Kinsel, M. E. Preparation of zeolite UTD-1 films by pulsed laser ablation: Evidence for oriented crystal growth. *Chem. Mater.* **10**, 464–466 (1998).
85. Wang, Z. B. & Yan, Y. S. Controlling crystal orientation in zeolite MFI thin films by direct *in situ* crystallization. *Chem. Mater.* **13**, 1101–1107 (2001).
86. Xomeritakis, G., Lai, Z. P. & Tsapatsis, M. Separation of xylene isomer vapors with oriented MFI membranes made by seeded growth. *Ind. Eng. Chem. Res.* **40**, 544–552 (2001).
87. Zhao, D. Y. *et al.* Continuous mesoporous silica films with highly ordered large pore structures. *Adv. Mater.* **10**, 1380–1385 (1998).
88. Yang, C. M. *et al.* Spin-on mesoporous silica films with ultralow dielectric constants, ordered pore structures, and hydrophobic surfaces. *Adv. Mater.* **13**, 1089–1102 (2001).
89. Liu, J. *et al.* Mesoporous silica film from a solution containing a surfactant and methods of making same. US Patent 6,329,017 (2001).
90. Wirsberger, G., Scott, B. J. & Stucky, G. D. pH sensing with mesoporous thin films. *Chem. Commun.* 119–120 (2001).
91. Fan, H. Y. *et al.* Rapid prototyping of patterned functional nanostructures. *Nature* **405**, 56–60 (2000).
92. Lu, Y. F. *et al.* Evaporation-induced self-assembly of hybrid bridged silsesquioxane film and particulate mesophases with internal organic functionality. *J. Am. Chem. Soc.* **122**, 5258–5261 (2000).
93. Baskaran, S. *et al.* Low dielectric constant mesoporous silica films through molecularly templated synthesis. *Adv. Mater.* **12**, 291–294 (2000).
94. Doshi, D. A. *et al.* Optically defined multifunctional patterning of photosensitive thin-film silica mesophases. *Science* **290**, 107–111 (2000).
95. Tolbert, S. H., Firouzi, A., Stucky, G. D. & Chmelka, B. F. Magnetic field alignment of ordered silicate-surfactant composites and mesoporous silica. *Science* **278**, 264–268 (1997).
96. Hillhouse, H. W., Okubo, T., van Egmond, J. W. & Tsapatsis, M. Preparation of supported mesoporous silica layers in a continuous flow cell. *Chem. Mater.* **9**, 1505–1507 (1997).
97. Grün, M., Kurganov, A. A., Schacht, S., Schüth, F. & Unger, K. K. Comparison of an ordered mesoporous aluminosilicate, silica, alumina, titania and zirconia in normal-phase high-performance liquid chromatography. *J. Chromatogr. A* **740**, 1–9 (1996).
98. Thoenen, C., van de Walle, K., Vankelecom, I. F. J. & Jacobs, P. J. The use of M41S materials in chiral HPLC. *Chem. Commun.* 1841–1842 (1999).
99. Sierra, L., Lopez, B., Ramirez, A. & Guth, J. L. Evaluation of mesoporous silicas as stationary phases for high performance liquid chromatography (HPLC). *Stud. Surf. Sci. Catal.* **135**, [CD ROM] Paper 18-P-06 (Elsevier, 2001).
100. Miller, R. D. In search of low- k dielectrics. *Science* **286**, 421–423 (1999).
101. McCoy, M. Completing the circuit. *Chem. Eng. News* **78**, 13–24 (2000).
102. Wang, Z., Wang, H., Mitra, A., Huang, L. & Yan, Y. Pure-silica zeolite low- k dielectric thin film by spin-on process. *Stud. Surf. Sci. Catal.* **135**, [CD-ROM] Paper 20-P-11 (Elsevier, 2001).
103. Lauffer, R. B. Paramagnetic metal complexes as water protein relaxation agents for NMR imaging: Theory and design. *Chem. Rev.* **87**, 901–927 (1987).
104. Cacheris, W. P., Quay, S. C. & Rocklage, S. M. The relationship between thermodynamics and the toxicity of gadolinium complexes. *Magn. Reson. Imaging* **8**, 467–481 (1990).
105. Balkus, K. J. Jr, Sherry, A. D. & Young, S. W. Zeolite-enclosed transition and rare earth metal ions as contrast agents for the gastrointestinal tract. US Patent 5,122,363 (1992).
106. Balkus, K. J. Jr, Bresnaka, I., Kowalak, S. & Young, S. W. The application of molecular sieves as magnetic resonance image contrast agents. *Mater. Res. Soc. Symp. Ser. Proc.* **223**, 225–230 (1991).
107. Balkus, K. J. Jr & Shi, J. Studies of gadolinium (III)-modified hectorite clays as potential oral MRI contrast agents. *J. Phys. Chem.* **100**, 16429–16434 (1996).
108. Balkus, K. J. Jr & Shi, J. A study of suspended agents for gadolinium (III)-exchanged heterite. An oral magnetic resonance imaging contrast agent. *Langmuir* **12**, 6277–6281 (1996).
109. Vietze, U. *et al.* Zeolite-dye microlasers. *Phys. Rev. Lett.* **81**, 4628–4631 (1998).
110. Ihlein, G., Schüth, F., Krauss, O., Vietze, U. & Laeri, F. Alignment of a laser dye in the channels of the AlPO_4 -5 molecular sieve. *Adv. Mater.* **10**, 1117–1119 (1998).
111. Weiß, O., Schüth, F., Benmohammadi, L. & Laeri, F. Potential microlasers based on AlPO_4 -5/DCM composites. *Stud. Surf. Sci. Catal.* **135**, [CD-ROM] Paper 21-O-04 (Elsevier, 2001).
112. Yang, P. *et al.* Mirrorless lasing from mesostructured wave guides patterned by soft lithography. *Science* **287**, 465–467 (2000).
113. Wirsberger, G. & Stucky, G. D. Microring lasing from dye-doped silica/block copolymer nanocomposites. *Chem. Mater.* **12**, 2525–2527 (2000).
114. Wada, Y. *et al.* High efficiency near-IR emission of Nd(III) based low-vibrational environment in cages of nanosized zeolites. *J. Am. Chem. Soc.* **122**, 8583–8584 (2000).
115. Enzel, P. & Bein, T. Poly(acrylonitrile) chains in zeolite channels—Polymerization and pyrolysis. *Chem. Mater.* **4**, 819–824 (1992).
116. Kyotani, T., Nagai, T., Inoue, S. & Tomita, A. Formation of new type of porous carbon by carbonization in zeolite nanochannels. *Chem. Mater.* **9**, 609–615 (1997).
117. Ma, Z., Kyotani, T., Liu, Z., Terasaki, O. & Tomita, A. Very high surface area microporous carbon with a three-dimensional nano-array structure: synthesis and its molecular structure. *Chem. Mater.* **13**, 4413–4415 (2001).
118. Johnson, S. A., Brigham, E. S., Olliver, P. J. & Mallouk, T. E. Effect of micropore topology on the structure and properties of zeolite polymer replicas. *Chem. Mater.* **9**, 2448–2458 (1997).
119. Ryoo, R., Joo, S. H. & Jun, J. Synthesis of highly ordered carbon molecular sieves via template-

- mediated structural transformation. *J. Phys. Chem. B* **103**, 7743–7746 (1999).
120. Lee, J. W., Yoon, S. H., Hyeon, T. H., Oh, S. M. & Kim, K. B. Synthesis of new mesoporous carbon and its application to electrochemical double-layer capacitors. *Chem. Commun.* 2177–2178 (1999).
121. Kruk, M., Jaroniec, M., Ryoo, R. & Joo, S. H. Characterization of ordered mesoporous carbons synthesized using MCM-48 silicas as templates. *J. Phys. Chem. B* **104**, 7960–7968 (2000).
122. Yoon, S. B., Kim, J. Y. & Yu, J. S. Synthesis of highly ordered nanoporous carbon molecular sieves from silylated MCM-48 using divinyl benzene as precursor. *Chem. Commun.* 559–560 (2001).
123. Joo, S. H. *et al.* Ordered nanoporous arrays of carbon supporting high dispersions of platinum nanoparticles. *Nature* **412**, 169–172 (2001).
124. Wang, N., Tang, Z. K., Li, G. D. & Chen, J. S. Single-walled 4 Å carbon nanotube arrays. *Nature* **408**, 50–51 (2000).
125. Tang, Z. K. *et al.* Superconductivity in 4 Å single-walled carbon nanotubes. *Science* **292**, 2462–2465 (2001).
126. Dagani, R. Littlest carbon nanotube debuts. *Chem. Eng. News* **78**, 9–10 (2000).
127. Wilson, E. Superconducting nanotubes. *Chem. Eng. News* **79**, 8 (2001).
128. Davis, M. E. Evolution of extra-large pore materials. *Stud. Surf. Sci. Catal.* **135**, 29–36 (2001).
129. Bull, I. *et al.* Imposition of polarity on a centrosymmetric zeolite host: The effect of fluoride ions on template ordering in zeolite IFR. *J. Am. Chem. Soc.* **122**, 7128–7129 (2000).
130. Corma, A., Nemeth, L. T., Renz, M. & Valencia, S. Sn-zeolite beta as a heterogeneous chemoselective catalyst for Baeyer-Villiger oxidations. *Nature* **412**, 423–425 (2001).
131. Dessau, R. M., Schlenker, J. L. & Higgins, J. B. Framework topology of AlPO_4 -8—The first 14-ring molecular-sieve. *Zeolites* **10**, 522–524 (1990).
132. Vogt, E. T. C. & Richardson, J. W. The reversible transition of the molecular-sieve VPI-5 into AlPO_4 -8 and the structure of AlPO_4 -8. *J. Solid State Chem.* **87**, 469–471 (1990).
133. Loiseau, T. & Ferey, G. Oxyfluorinated microporous compounds. 7. Synthesis and crystal structure of ULM-5, a new fluorinated gallophosphate $\text{Ga}_{16}(\text{PO}_4)_{14}(\text{HPO}_4)_2(\text{OH})_2\text{F}_7 \cdot [\text{H}_3\text{N}(\text{CH}_2)_6\text{NH}_3]_4 \cdot 6 \text{H}_2\text{O}$ with 16-membered rings and both bonding and encapsulated F^- . *J. Solid State Chem.* **111**, 403–415 (1994).
134. Loiseau, T. & Ferey, G. Synthesis and crystal structure of ULM-16, a new open framework fluorinated gallium phosphate with 16-ring channels. *Mater. Res. Soc. Symp. Ser. Proc.* **431**, 27–38 (1996).
135. Zhou, Y. *et al.* A large 24-membered-ring germanate zeolite-type open-framework structure with three-dimensional intersecting channels. *Angew. Chem. Int. Edn Engl.* **40**, 2166–2168 (2001).
136. Lin, C. H., Wang, S. L. & Lii, K. H. $[\text{Ga}_2(\text{DETA})(\text{PO}_4)_2] \cdot (2\text{H}_2\text{O})$ (DETA = diethylenetriamine): a novel porous gallium phosphate containing 24-ring channels. *J. Am. Chem. Soc.* **123**, 4649–4650 (2001).
137. Rohrig, C. & Gies, H. A new zincosilicate zeolite with 9-ring channels. *Angew. Chem. Int. Edn Engl.* **34**, 63–65 (1995).
138. McCusker, L., Grosse-Kunstleve, R. W., Baerlocher, Ch., Yoshikawa, M. & Davis, M. E. Synthesis optimization and structure analysis of the zincosilicate molecular sieve VPI-9. *Microporous Mater.* **6**, 295–309 (1996).
139. Grosse-Kunstleve, R. W. *Zeolite Structure Determination From Powder Data: Computer-based Incorporation of Crystal Chemical Information*. PhD thesis, Swiss Federal Inst. Technol., Zürich (1996).
140. Park, S. H., Daniels, P. & Gies, H. RUB-23: A new microporous lithosilicate containing spiro-5 building units. *Microporous Mesoporous Mater.* **37**, 129–143 (2000).
141. Park, S. H., Parise, J. B. & Gies, H. Optimized synthesis and structural properties of lithosilicate RUB-29. *Stud. Surf. Sci. Catal.* **135**, [CD-ROM] Paper 09-O-05 (Elsevier, 2001).
142. Li, H., Eddaoudi, M., Plevet, J., O’Keeffe, M. & Yaghi, O. M. $\text{Ge}_2\text{ZrO}_6\text{F}_2 \cdot (\text{H}_2\text{DAB})\text{H}_2$: A 4-connected microporous material with “bowtie” building units and an exceptional proportion of 3-rings. *J. Am. Chem. Soc.* **122**, 12409–12410 (2000).

Correspondence and requests for materials should be addressed to the author (e-mail: mdavis@cheme.caltech.edu).

Angiotensin-converting enzyme 2 is an essential regulator of heart function

Michael A. Crackower^{*†‡}, Renu Sarao^{*§||}, Gavin Y. Oudit^{‡||¶#}, Chana Yagil[☆], Ivona Kozieradzki^{*§}, Sam E. Scanga^{**}, Antonio J. Oliveira-dos-Santos^{*}, Joan da Costa^{*}, Liyong Zhang^{*†‡§}, York Pei[#], James Scholey[#], Carlos M. Ferrario^{††}, Armen S. Manoukian^{**}, Mark C. Chappell^{†‡}, Peter H. Backx^{‡¶#}, Yoram Yagil[☆] & Josef M. Penninger^{*†‡§}

^{*} Amgen Research Institute/Ontario Cancer Institute; and [†] Departments of Medical Biophysics and Immunology; and [‡] The Heart & Stroke/Richard Lewar Centre for Excellence in Cardiovascular Research; and [¶] Department of Physiology; and [#] Department of Medicine and the University Health Network; and ^{**} Division of Cellular & Molecular Biology, University Health Network, Ontario Cancer Institute, University of Toronto, University Avenue, Toronto, Ontario M5G 2M9, Canada

[§] IMBA, Institute for Molecular Biotechnology of the Austrian Academy of Sciences, c/o Dr Bohr Gasse, A-1030, Vienna, Austria

[☆] Laboratory for Molecular Medicine, Department of Nephrology and Hypertension, Faculty of Health Science, Ben-Gurion University Barzilai Medical Center Campus, Ashkelon 78306, Israel

^{††} The Hypertension and Vascular Disease Center, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157-1032, USA

^{||} These authors contributed equally to the work

Cardiovascular diseases are predicted to be the most common cause of death worldwide by 2020. Here we show that angiotensin-converting enzyme 2 (*ace2*) maps to a defined quantitative trait locus (QTL) on the X chromosome in three different rat models of hypertension. In all hypertensive rat strains, *ACE2* messenger RNA and protein expression were markedly reduced, suggesting that *ace2* is a candidate gene for this QTL. Targeted disruption of *ACE2* in mice results in a severe cardiac contractility defect, increased angiotensin II levels, and upregulation of hypoxia-induced genes in the heart. Genetic ablation of *ACE* on an *ACE2* mutant background completely rescues the cardiac phenotype. But disruption of *ACER*, a *Drosophila* *ACE2* homologue, results in a severe defect of heart morphogenesis. These genetic data for *ACE2* show that it is an essential regulator of heart function *in vivo*.

Cardiovascular disease will be the greatest health care burden of the twenty-first century¹. A major risk factor for heart disease is high blood pressure². Hypertension is a multifactorial quantitative trait controlled by both genetic and environmental factors³. While much is known about environmental factors that can contribute to high blood pressure, such as diet and physical activity, less is known about the genetic factors that are responsible for predisposition to cardiovascular disease. Despite the identification of several putative

genetic quantitative trait loci (QTL) associated with hypertension in animal models⁴, none of these loci have been translated into genes⁵. Thus, the molecular and genetic mechanisms underlying hypertension and other cardiovascular diseases remain obscure.

One important regulator of blood pressure homeostasis is the renin-angiotensin system (RAS). The protease renin cleaves angiotensinogen into the inactive decameric peptide angiotensin I (AngI). Angiotensin-converting enzyme (ACE) then catalyses the

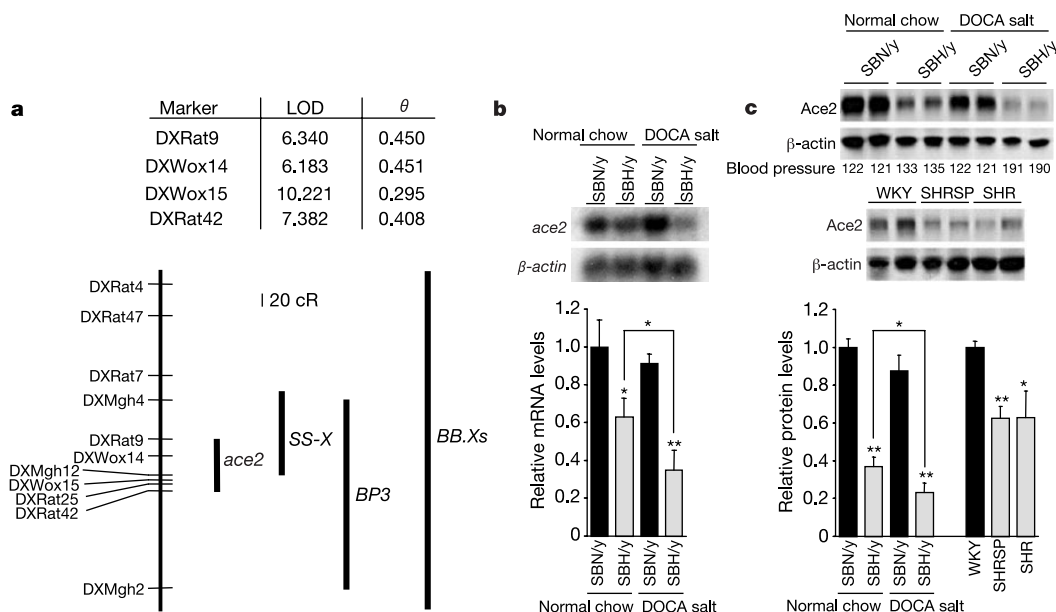


Figure 1 Association of *ACE2* and hypertension in the rat. **a**, Results of radiation hybrid mapping of rat *ace2*, compared to the mapping of a QTL identified in Sabra salt-sensitive animals (SS-X), SHRSP (BP3), and SHR rats (BB.Xs). Polymorphic marker names are indicated to the left of the ideogram. LOD scores and theta values for markers linked to *ace2* are shown. cR, centiradians. **b**, Northern blot analysis of *ace2* mRNA from kidneys of

Sabra SBH/y and SBN/y rats. **c**, Western blot analysis of *ACE2* protein levels from kidneys of Sabra SBH/y and their control SBN/y rats, as well as SHR and SHRSP and their control WKY rats. Systolic blood pressure in mm Hg for the respective Sabra rats is indicated. Bars show mean values $\pm$ s.e.m. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$ ($n = 4$, for all groups).

cleavage of the AngI into the active octomer angiotensin II (AngII), which can contribute to hypertension by promoting vascular smooth muscle vasoconstriction and renal tubule sodium reabsorption^{6,7}. ACE mutant mice display spontaneous hypotension, partial male infertility, and kidney malformations^{8,9}. In humans, an ACE polymorphism has been associated with determinants of renal and cardiovascular function¹⁰, and pharmacological inhibition of ACE and AngII receptors are effective in lowering blood pressure¹¹ and preventing kidney disease¹². In addition, inhibition of ACE and AngII receptors has beneficial effects in heart failure¹³.

Recently, a homologue of ACE, termed ACE2, has been identified; it is predominantly expressed in the vascular endothelial cells of the kidney and heart^{14,15}. Unlike ACE, ACE2 functions as a carboxypeptidase, cleaving a single residue from AngI, generating Ang1-9 (refs 14, 15), and a single residue from AngII to generate Ang1-7 (ref. 14). These *in vitro* biochemical data suggest that ACE2 may modulate the RAS and thus affect blood pressure regulation. In addition, it has been shown that ACE2 can cleave other peptide substrates^{14,15}. Two different ACE2 homologues have been identified in flies^{16,17}. Nevertheless, the *in vivo* role of ACE2 in the cardiovascular system and the RAS is not known.

ACE2 maps to a QTL on the X chromosome in hypertensive rats

Hypertension and most cardiovascular diseases are multifactorial in nature and disease pathogenesis is influenced by multiple genetic susceptibility loci³. In various recombinant rat models, several QTL for hypertension have been identified. *ace2* maps to the X chromosome in human¹⁴ and a QTL has been mapped to the X chromosome in several rat models of hypertension with no candidate gene ascribed to it as yet^{18–20}. Angiotensin receptor II, which also maps to the X chromosome, has previously been excluded from this candidate interval²¹.

We speculated that *ace2* could be a candidate gene for this QTL. Radiation hybrid mapping showed that the rat *ace2* gene maps on

the X chromosome with significant logarithm-of-odds (LOD) scores to markers *DXRat9*, *DXWox14*, *DXWox15* and *DXRat42*, placing *ace2* between *DXRat9* and *DXRat42* (Fig. 1a). Comparative mapping showed that the *ace2* map position overlaps with a QTL interval for hypertension identified in Sabra salt-sensitive rats found between markers *DXMgh12* and *DXRat8* (SS-X) (ref. 18). Moreover, the chromosomal *ace2* region maps to the BP3 QTL interval defined in stroke-prone spontaneously hypertensive rats (SHRSP) rats¹⁹, and the hypertensive *BB.Xs* QTL identified on the X chromosome of spontaneous hypertensive rats (SHR) by congenic analysis²⁰ (Fig. 1a). Thus, *ace2* maps to a QTL on the rat X chromosome that has been identified in three separate models of spontaneous and diet-induced hypertension.

Downregulation of ACE2 expression in hypertensive rats

Because the kidney is a major site of blood pressure regulation²², we determined *ace2* expression levels in the kidneys of these three hypertensive rat strains. We initially measured *ace2* mRNA levels in the kidneys of salt-sensitive Sabra hypertensive (SBH/y) rats and control salt-resistant Sabra normotensive (SBN/y) rats. Salt loading (with deoxycorticosterone acetate (DOCA) salt) had no effect on *ace2* mRNA expression in normotensive SBN/y rats. In SBH/y rats, salt loading and the development of hypertension were associated with a significant reduction in *ace2* mRNA expression compared to normotensive SBN/y rats (Fig. 1b). *ace2* mRNA was also lower in SBH/y rats fed a regular diet ('normal chow') when compared to SBN/y control rats fed a similar diet. This latter finding suggests that downregulation of ACE2 may be independent of blood pressure. However, the possibility that ACE2 expression is controlled by blood pressure cannot be excluded.

To measure ACE2 protein levels, we generated ACE2 (amino acids 206–225 of mouse ACE2) specific rabbit antiserum, which cross-reacts with both rat and human ACE2 (not shown). In line with the decreased ACE2 mRNA expression, ACE2 protein expression was markedly reduced in SBH/y animals that were fed a normal diet (Fig. 1c). Increase in blood pressure of SBH/y rats following a 4-week diet of DOCA salt correlated with a further decrease in ACE2 protein expression (Fig. 1c). Salt loading did not trigger increased blood pressure²³, nor did it alter ACE2 expression in salt-resistant SBN/y control rats (Fig. 1c). ACE2 protein levels were also significantly decreased in the kidneys of SHRSP and SHR rats as compared to their WKY controls (Fig. 1c). Moreover, the levels of ACE2 mRNA were markedly reduced in SHRSP and SHR rats (not shown). Cloning and sequencing of the coding region of *ace2* in the hypertensive rat strains did not reveal any sequence changes, indicating that reduced ACE2 expression probably results from polymorphisms that control *ace2* gene expression. The map position and reduced expression of ACE2 in three different rat strains indicate that *ace2* is a strong candidate gene for this hypertensive QTL on the X chromosome.

Generation of ACE2 knockout mice

To validate the candidacy of *ace2* as a QTL and to test whether ACE2 has indeed an essential role in cardiovascular physiology and the pathogenesis of cardiovascular disease, we disrupted the *ace2* gene in mouse by homologous recombination (see Supplementary Information). The null mutation of *ace2* was verified by the absence of *ace2* mRNA transcripts and protein by northern (not shown) and western blot analyses (Fig. 2a). *ace* mRNA expression in the kidneys and hearts was not altered in *ace2* mutant mice (Fig. 2b). ACE2 null mice were born at the expected mendelian frequency, appeared healthy, and did not display any gross detectable alterations in all organs analysed. Moreover, in contrast to *ace*^{−/−} male mice that display significantly reduced fertility, both male and female *ace2* null mice are fertile.

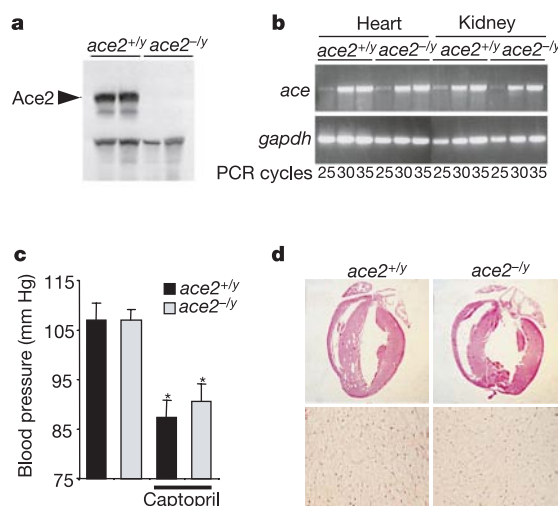


Figure 2 ACE2-deficient mice. **a**, Western blot analysis of ACE2 protein expression in the kidneys of *ace2*^{+/y} and *ace2*^{−/y} mice. **b**, Polymerase chain reaction with reverse transcription of RNA (RT-PCR) analysis of ACE mRNA expression in the heart and kidneys of *ace2*^{+/y} and *ace2*^{−/y} mice. **c**, Blood pressure measurements in 3-month-old *ace2*^{+/y} (*n* = 8) and *ace2*^{−/y} (*n* = 8) mice in the absence (left panels) or presence of the ACE blocker captopril. Mean values ± s.e.m. are shown (double asterisk; *P* < 0.01). **d**, The top panels show haematoxylin and eosin stained sections of hearts isolated from 6-month-old *ace2*^{+/y} and *ace2*^{−/y} mice. We note the enlarged left and right ventricles in *ace2*^{−/y} mice. However, the overall heart size is comparable between both genotypes and there is no evidence for cardiac hypertrophy macroscopically or in isolated cardiomyocytes. The bottom panels show an absence of interstitial fibrosis in *ace2*^{−/y} mice. As shown by PSR staining (see Methods).

Normal blood pressure in *ace2* mutant mice

It has been previously shown that *ace* mutant mice display reduced blood pressure and kidney pathology^{8,9}. Therefore, we first tested whether loss of ACE2 expression affects blood pressure homeostasis and/or kidney development or function. Loss of ACE2 did not result in alteration of blood pressure in 3-month-old *ace2*^{-/-} male (Fig. 2c) or *ace2*^{-/-} female mice (not shown) as compared to their control littermates. Because it was possible that ACE could compensate for the loss of ACE2, we treated *ace2*-deficient mice with captopril, which blocks ACE but not ACE2 function^{14,15}. However, *in vivo* inhibition of ACE with captopril reduced the blood pressure of *ace*^{-/-} male mice to a similar extent as was observed in captopril-treated wild-type littermates (Fig. 2c). Thus, even in a scenario of ACE inhibition, loss of ACE2 has no apparent direct effect on blood pressure homeostasis. In addition, no alterations in kidney ultrastructure or function or anaemia could be detected (not shown).

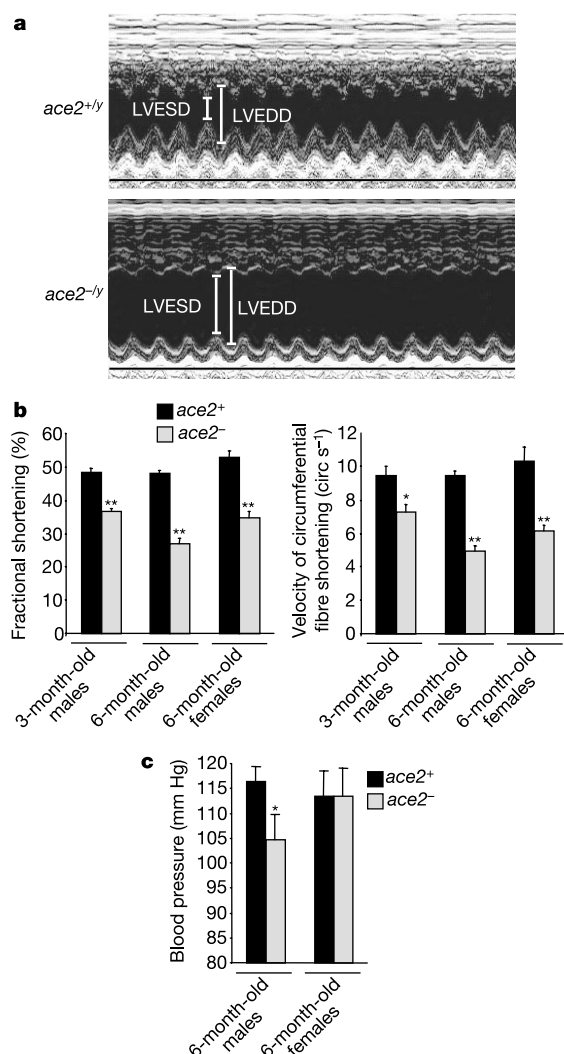


Figure 3 Loss of ACE2 results in severe contractile dysfunction. **a**, Echocardiographic measurements of contracting hearts in a 6-month-old *ace2*^{+/-} mouse and two *ace2*^{-/-} mice. Arrows indicate the distance between systolic contraction (LVESD) and diastolic relaxation (LVEDD). **b**, Per cent fractional shortening and velocity of circumferential fibre shortening (circumferences s⁻¹) in 6-month-old *ace2*^{+/-} (*n* = 8) and *ace2*^{-/-} (*n* = 8) mice and 6-month-old *ace2*^{+/-} (*n* = 5) and *ace2*^{-/-} (*n* = 5) female mice. **c**, Tail-cuff blood pressure measurements in 6-month-old male *ace2*^{+/-} (*n* = 8) and *ace2*^{-/-} (*n* = 8) mice and 6-month-old female *ace2*^{+/-} (*n* = 5) and *ace2*^{-/-} (*n* = 5) mice. Mean values ± s.e.m. are shown. Asterisk, *P* < 0.05 and double asterisk, *P* < 0.01.

Loss of ACE2 impairs heart function

Pharmacological inhibition of ACE or AngII receptors suggested a role for the RAS in the regulation of heart function and cardiac hypertrophy¹³. However, neither *ace*^{8,9} nor *angiotensinogen*²⁴ null mice develop any overt heart disease. ACE2 is highly expressed in the vasculature of the heart, so we analysed hearts of *ace2*-deficient mice. Hearts of *ace2* mutant mice display a slight wall thinning of the left ventricle and increased chamber dimensions (Fig. 2d). Thinning of the anterior left ventricular wall and increase in the left ventricle end diastolic dimension in *ace2*-deficient hearts can be also seen by echocardiography (Table 1). These structural changes are primarily observed in 6-month-old male mice. However, heart weights and heart to body weight ratios were comparable between age-matched 3-month-old (not shown) and 6-month-old *ace2*^{-/-} and *ace2*^{+/-} mice (see Supplementary Information). Echocardiography also showed that the left ventricular-mass and its ratio to body weight were normal (Table 1). We also failed to observe structural and biochemical changes characteristic of dilated cardiomyopathy as there was no indication of interstitial cardiac fibrosis (Fig. 2d) nor prototypical changes in ANF, BNP, α-MHC, β-MHC and skeletal muscle actin gene expression (not shown). In addition, individual cardiomyocytes of ACE2 null mice exhibited no evidence of hypertrophy and we did not observe any evidence of altered cardiomyocyte apoptosis in ACE2 null mice as detected by TdT-dependent dUTP-biotin nick end labelling (TUNEL) staining (not shown). Thus, despite mild dilation of hearts in 6-month-old ACE2 null mice, there was no evidence of cardiac hypertrophy or dilated cardiomyopathy.

Assessment of cardiac function by echocardiography revealed that all *ace2*^{-/-} male and *ace2*^{-/-} female mice exhibit severe reduction in cardiac contractility as determined by decreased left ventricle fractional shortening, and decreased velocity of circumferential fibre shortening (Table 1 and Fig. 3a, b). The decrease in function was found to be more severe in 6-month-old male than in age-matched female mice. In addition, 3-month-old male mice had a less pronounced phenotype than older animals (Table 1 and Fig. 3b), suggesting a progression in the phenotype (Table 1). Consistent with the decreased cardiac contractility, 6-month-old

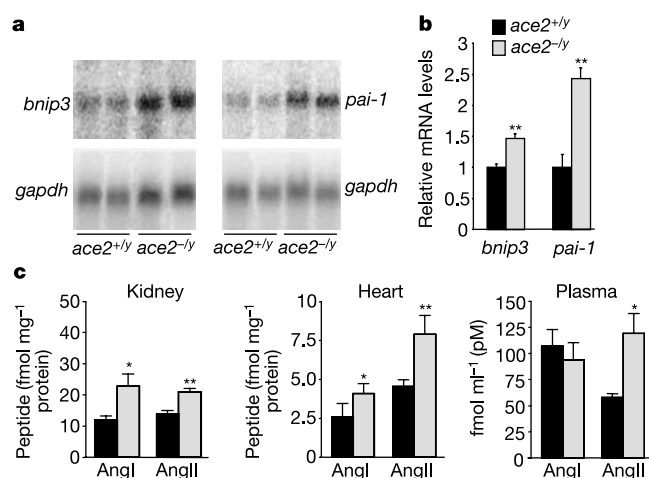


Figure 4 Upregulation of hypoxia markers and increased angiotensin II levels in the absence of ACE2. **a**, **b**, Northern blot analysis of BNIP3 and PAI-1 mRNA expression levels, two hypoxia-inducible genes in 6-month-old *ace2*^{+/-} and *ace2*^{-/-} male mice. **a**, Individual northern blot data; **b**, relative levels of BNIP3 and PAI-1 mRNA levels normalized to the gapdh control. Double asterisk, *P* < 0.01 (*n* = 5). **c**, Angiotensin I (AngI) and Angiotensin II (AngII) peptide levels in the heart, kidneys and plasma of male *ace2*^{+/-} (*n* = 8) and *ace2*^{-/-} (*n* = 8) littermate mice. Peptide levels were determined by radioimmunoassays. Mean peptide levels ± s.e.m. are shown. Double asterisk, *P* < 0.01.

male *ace2*^{-/-} mice exhibited reduced blood pressure (Fig. 3c), a feature not found in age-matched *ace2*^{-/-} females and 3-month-old males. This suggests that the reduction in blood pressure may be the result of severe cardiac dysfunction and not a direct effect of loss of ACE2 on systemic blood pressure. To confirm the echocardiographic defects in cardiac function, invasive haemodynamic measurements were performed in *ace2* null mice. These invasive haemodynamic measurements showed that both the maximum and minimum rates of change of left ventricular pressure, (dP/dt)_{max} and (dP/dt)_{min}, were markedly reduced in the *ace2* mutant mice (Table 1), indicating severe impairment of contractile heart function. Loss of ACE2 also resulted in a significant decrease in aortic and ventricular pressures consistent with the observed reductions in cardiac contractility (data not shown). These results show that ACE2 is an important regulator of heart function *in vivo*.

Upregulation of hypoxia-inducible genes in *ace2* null mice

The severe contractile dysfunction and mild dilation in the absence of hypertrophy or cardiac fibrosis in *ace2* null mice resembles cardiac stunning/hibernation in human and animal models^{25,26}. Cardiac stunning and hibernation are adaptive responses to chronic hypoxia such as in coronary artery disease or following bypass surgery²⁷. Because ACE2 is highly expressed in vascular endothelial cells but contractility is controlled by cardiomyocytes, we speculated that loss of ACE2 could result in cardiac hypoxia. We therefore analysed changes in the expression levels of hypoxia-inducible genes such as *BNIP3* (ref. 28) and *PAI-1* (ref. 29) by northern blotting. In the hearts of all *ace2* null mice analysed, mRNA expression of BNIP3 and PAI-1 was markedly upregulated compared to their wild-type littermates (Fig. 4a and b). The magnitude of increased expression of these markers of hypoxia is similar to that previously observed in other hypoxic models such as in myocyte-specific vascular endothelial growth factor mutant mice³⁰. Thus, loss of ACE2 results in the induction of a hypoxia-regulated gene expression profile.

Increased angiotensin II levels in *ace2* null mice

Because ACE2 functions as a carboxypeptidase, cleaving a single residue from AngI to generate Ang1-9 (refs 14, 15) and a single residue from AngII to generate Ang1-7 (ref. 14), we hypothesized that ACE2 may regulate the RAS by competing with ACE for the substrate AngI and/or cleaving and inactivating AngII. If correct, loss of ACE2 should increase AngII levels *in vivo*. Using radioimmunoassays, AngII levels were indeed found to be significantly increased in the kidneys, hearts and plasma of *ace2* mutant mice (Fig. 4c). In addition, an increase in AngI was observed in the heart and kidney of *ace2* mutant mice (Fig. 4c), consistent with AngI

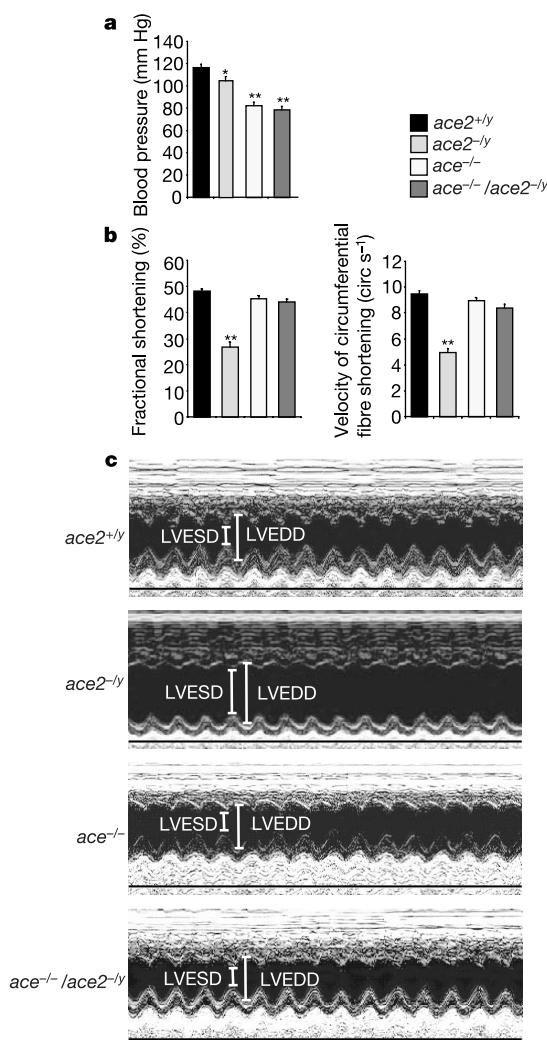


Figure 5 ACE/ACE2 double mutant mice do not develop cardiac dysfunction. **a**, Blood pressure measurements in 6-month-old *ace2*^{+/y} (*n* = 8), *ace2*^{-/y} (*n* = 8), *ace*^{-/-} (*n* = 8), and *ace*^{-/-}/*ace2*^{-/y} double mutant (*n* = 6) male mice. **b**, Per cent fractional shortening and velocity of circumferential fibre shortening in 6-month-old male *ace2*^{+/y} (*n* = 8), *ace2*^{-/y} (*n* = 8), *ace*^{-/-} (*n* = 8) and *ace*^{-/-}/*ace2*^{-/y} double mutant (*n* = 6) littermates. **c**, Echocardiographic measurements of contracting hearts in 6-month-old male *ace2*^{+/y}, *ace2*^{-/y}, *ace*^{-/-} and *ace*^{-/-}/*ace2*^{-/y} double mutant littermate mice. Mean values ± s.e.m. are shown. Asterisk, *P* < 0.05 and double asterisk, *P* < 0.01.

Table 1 Heart functions of *ace2* null mice

	3-month-old males		6-month-old males		6-month-old females	
	<i>ace2</i> ^{+/y}	<i>ace2</i> ^{-/y}	<i>ace2</i> ^{+/y}	<i>ace2</i> ^{-/y}	<i>ace2</i> ^{+/y}	<i>ace2</i> ^{-/-}
	<i>n</i> = 7	<i>n</i> = 7	<i>n</i> = 8	<i>n</i> = 8	<i>n</i> = 3	<i>n</i> = 3
Heart rate (b.p.m.)	469 ± 12	466 ± 18	495 ± 15	482 ± 12	460 ± 6	452 ± 14
Anterior left ventricular wall thickness (mm)	0.65 ± 0.02	0.62 ± 0.01	0.66 ± 0.01	0.59 ± 0.02*	0.65 ± 0.02	0.57 ± 0.04
LVEDD (mm)	4.09 ± 0.04	4.25 ± 0.10	4.12 ± 0.10	4.49 ± 0.12*	3.71 ± 0.10	1.11 ± 0.07
LVESD (mm)	2.11 ± 0.04	2.69 ± 0.09**	2.13 ± 0.06	3.29 ± 0.14**	1.74 ± 0.07	2.68 ± 0.11**
Fractional shortening (%)	48.42 ± 1.14	36.73 ± 0.89**	48.12 ± 0.84	26.76 ± 1.78**	53.00 ± 1.68	34.85 ± 1.76**
Peak aortic outflow velocity (m s ⁻¹)	0.992 ± 0.029	0.922 ± 0.033	0.902 ± 0.044	0.802 ± 0.037*	0.875 ± 0.048	0.809 ± 0.044
Velocity of circumferential fibre shortening (circ s ⁻¹)	9.49 ± 0.48	7.28 ± 0.45*	9.46 ± 0.26	4.93 ± 0.33**	10.34 ± 0.81	6.16 ± 0.31**
Calculated left ventricular mass (mg)	95.69 ± 1.88	95.50 ± 2.28	102.41 ± 4.06	100.49 ± 3.39	92.27 ± 2.02	89.60 ± 1.54
Left ventricular mass/body mass (mg g ⁻¹)	3.32 ± 0.08	3.53 ± 0.18	3.19 ± 0.10	3.32 ± 0.15	3.29 ± 0.37	3.30 ± 0.09
(dP/dt) _{max}	n.d.	n.d.	5,579 ± 422	3,034 ± 124**	n.d.	n.d.
(dP/dt) _{min}	n.d.	n.d.	-5,055 ± 257	-2,929 ± 271**	n.d.	n.d.

**P* < 0.05 *ace2*^{-/y} versus *ace2*^{+/y} or *ace2*^{-/-} versus *ace2*^{+/y}.

***P* < 0.01 *ace2*^{-/y} versus *ace2*^{+/y} or *ace2*^{-/-} versus *ace2*^{+/y}.

b.p.m., heart beats per minute; LVEDD, left ventricle end diastolic dimension; LVESD, left ventricle end systolic dimension; (dP/dt)_{max}, maximum first derivative of the change in left ventricular pressure with time; (dP/dt)_{min}, minimum first derivative of the change in left ventricular pressure with time. n.d. not determined.

being a substrate of ACE2 action *in vivo*. No differences in ACE mRNA levels were found in the hearts and kidneys of *ace2* mutant mice compared to controls, indicating that the increased AngII tissue levels were not due to increased ACE expression (Fig. 2b). These data show that ACE2 functions as a regulator of the RAS modulating endogenous levels of AngI and AngII.

Ablation of ACE expression rescues *ace2*-deficient mice

To determine whether genetic ablation of ACE in combination with disruption of ACE2 has any effect on the heart phenotype in *ace2* mutant mice, we generated *ace/ace2* double mutant mice. These double mutant mice were born at the expected mendelian ratio and appear healthy. Blood pressure (Fig. 5a) and kidney defects (not shown) in the *ace/ace2* double null mice were similar to that of *ace* single mutant mice. Fertility of the *ace/ace2* double mutant mice was not addressed. Thus, loss of both ACE and ACE2 does not cause any apparent disease in addition to that seen in *ace* single mutant mice.

The heart function of ACE knockout mice has not been previously reported, to our knowledge, so we first analysed the heart parameters in these mice. In *ace*^{-/-} mice, hearts are histologically normal (not shown) and no defect in heart function could be detected at 6 months of age (Fig. 5b, c). However, ablation of ACE expression on an *ace2* mutant background completely abolished the cardiac dysfunction phenotype of *ace2* single knockout mice (Fig. 5b, c). Using echocardiography, all heart functions of 6-month-old, age-matched *ace/ace2* double mutant mice were comparable to that of their *ace* single mutant and wild-type littermates (Fig. 5, Table 1 and Supplementary Information). Restoration of heart functions occurred in both male and female *ace/ace2* double mutant mice. Importantly, there was also no difference in blood pressure between *ace* and *ace/ace2* knockout mice (Fig. 5a), further implying that the reduced blood pressure in older male *ace2* mice is due to the dramatic decrease in heart function.

Heart defects in ACER mutant flies

ACE2 and ACE have critical roles in heart function in mice, so we analysed whether the role of ACE2 in the heart is evolutionarily conserved. Flies have two ACE2 homologues called ANCE¹⁶ and ACER¹⁷, both of which have a domain structure similar to ACE2 (see

Supplementary Information). Both fly homologues are expressed in developing heart cells of the *Drosophila* embryo^{16,17}. To determine what role these ACE homologues may have in heart development, we studied mutant flies that carry a P-element insertion in the *Acer* locus (P679) (ref. 31). Mutation of ACER results in early embryonic lethality (not shown). We, therefore, examined heart tube morphogenesis using two early markers (the even-skipped (Eve) and Tinman (Tin) proteins) for heart progenitor cells in *Drosophila* embryos.

Expression of Eve is the earliest available marker for heart progenitor cells^{32,33} and in wild-type stage-11 embryos, clusters of three Eve-positive heart progenitor cells form the earliest heart tube anlage (Fig. 6a). By contrast, ACER mutant flies display reduced numbers and disorganization of Eve positive progenitor cells (Fig. 6b). To study further whether ACER mutant mice have a defect in mesoderm that specifies heart progenitor cells, we used Tinman as a marker. Tinman provides the dorsal mesoderm with the competence to specify heart progenitors³⁴, a role conserved in *tinman*-related homeobox genes such as murine *Nkx2.5* that is essential for heart development³⁵. Similar to alterations in Eve expression, loss of ACER results in reduced numbers and disorganization of the Tinman positive dorsal mesoderm (Fig. 6c, d). Thus, ACER mutant flies have defective heart morphogenesis and ACER may have a role in the specification of heart progenitors in *Drosophila* embryos.

Discussion

Here we have shown that *ace2* maps to a QTL associated with hypertension in three rat models of high blood pressure. ACE2 levels are reduced in all of these hypertensive rat strains. In mice, genetic inactivation of ACE2 using homologous recombination results in increased AngII peptide levels, upregulation of hypoxia genes in the heart and severe cardiac dysfunction. Ablation of ACE expression on an *ace2*-deficient background completely abolished the heart-failure phenotype of *ace2* single knockout mice. Our results in *Drosophila* show that disruption of the fly ACE2 homologue, ACER¹⁷, results in a severe and lethal defect of heart morphogenesis. Our fly data also indicate that ACER has a role in the specification of heart progenitor cells. Thus, fly ACER and mammalian ACE/ACE2 may have divergent roles in heart progenitor cell specification versus the control of heart function.

Most cardiovascular diseases are multifactorial quantitative traits controlled by both genetic and environmental factors³. One major factor for cardiovascular disease is the RAS. The map position and reduced expression of ACE2 in three different rat strains indicate that *ace2* is a strong candidate gene for a hypertensive QTL on the X chromosome. However, loss of ACE2 in mice has no apparent effect on blood pressure even when ACE is inhibited. From previous studies in many systems, it has become clear that the interplay of several QTLs determine the phenotype. Currently it is thought that there are numerous genetic polymorphisms that serve to regulate blood pressure in humans and rodents³. Thus, altering one of the genes in the 'wrong' genetic mouse background would not be sufficient in itself to alter blood pressure. Only when a sufficient number of control mechanisms are altered in concert would high blood pressure ensue. But despite the lack of blood pressure changes in the ACE2 mutant mice, there is an increase in AngII levels. These data clearly indicate that ACE2 controls the levels of AngII *in vivo*, supporting the candidacy of *ace2* as a QTL that controls the RAS. Thus, the absence of increases in blood pressure in ACE2 mutant mice does not exclude *ace2* as a candidate gene for the QTL on the X chromosome. It will be interesting to determine whether single nucleotide polymorphisms in the human ACE2 locus correlate with changes in blood pressure.

Unexpectedly, loss of ACE2 in mice results in profound contractile dysfunction. The complete rescue of the heart phenotype in ACE/ACE2 double mutant mice indicates that ACE expression has a causative role in the onset of the heart dysfunction. Both ACE and

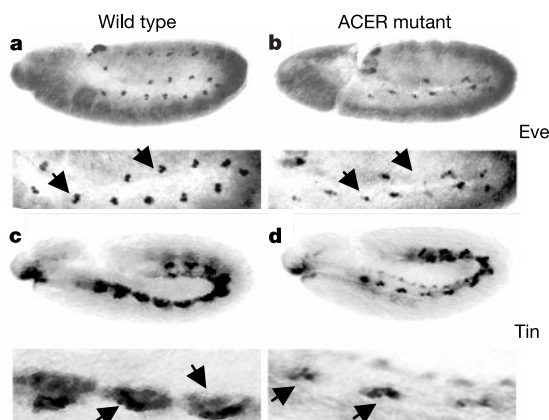


Figure 6 Expression of heart progenitor markers in *Drosophila* ACER mutant embryos. **a, b**, Expression of the heart progenitor cell marker even-skipped (Eve) in stage-11 wild-type and ACER mutant fly embryos. Bottom panels show close-ups of the Eve-positive heart progenitor cells. Arrows in **a** show clusters of three Eve-positive heart progenitor cells in wild-type embryos. In ACER mutant embryos, there is disorganization and reduction of Eve-positive cells (arrows). **c, d**, Expression of Tinman (Tin), a specification marker for heart progenitor cells, in stage-12 embryos. In ACER mutant embryos, the numbers of Tin-expressing cells is markedly reduced in the dorsal mesoderm as compared to wild-type *Drosophila* embryos.

ACE2 have been shown to be able to cleave AngI *in vitro*^{6,7,14,15}. Whereas ACE generates AngII from AngI, ACE2 can counteract the function of ACE by cleaving AngII or by competing with ACE for the substrate AngI. Thus, our mouse data suggest that increased AngII mediates the heart phenotype. Nevertheless, our data does not exclude the possibility that other metabolites^{6,7,14,15} of ACE2 and ACE contribute to the observed phenotypes. It will be important to assess directly the role of AngII in the heart defects of ACE2 mutant mice by pharmacological or genetic inhibition of AngII receptors, and determine if any other ACE2-modulating peptides are altered in these mice. We note that in human patients, inhibition of ACE or AngII receptors can improve the outcome of heart failure¹³. The contribution of ACE2 in human heart failure needs to be determined.

The heart phenotype found in ACE2 mutant mice is similar to that seen in cardiac stunning/hibernation such as in humans with coronary artery disease and cases of bypass surgery^{25–27}. In humans and in animal models of cardiac stunning/hibernation, chronic hypoxic conditions lead to compensatory changes in myocyte metabolism²⁷, upregulation of hypoxia-induced genes³⁰, and severe contractile dysfunction^{25–27,30}. Because ACE2 is expressed in the vascular endothelium and not in cardiac myocytes¹⁵, local increases in AngII may lead to vasoconstriction, resulting in hypoperfusion and hypoxia in the myocardium. It has been established that AngII can induce oxidative stress in endothelial cells^{36,37}, so the increase in AngII could also result in dysfunction of the vascular endothelium of the heart via the induction of oxidative stress. □

Methods

Cloning of mouse and rat ACE2 and chromosomal QTL mapping

Murine *ace2* was cloned from a proprietary EST database. Using a mouse *ace2* probe, we then screened a rat kidney cDNA (Invitrogen) to obtain a full-length rat cDNA as determined by DNA sequencing (see Supplementary Information). For chromosomal mapping, a rat *ace2* cDNA specific probe was used to screen a rat PAC library (RPCI-31, Research Genetics), identifying two positive clones (6M6 and 125K9). The end sequences of these clones were determined and rat-specific primers were designed (mc2L: 5'-TCAATTTACTGCTGAGGGGG-3'; and mc2R: 5'-GAGGGATAACCCAGTGCAAA-3') to determine the chromosomal map position of *ACE2* in rat by screening a radiation hybrid panel (RH07.5, Research Genetics). SHR and control Wistar Kyoto (WKY) rats were obtained from Harlan and maintained at the animal facilities of the Ontario Cancer Institute in accordance with institutional guidelines. Salt-resistant and salt-sensitive Sabra rats were bred and maintained at the animal facility of the Ben-Gurion University, Barzilai Medical Center, Israel. DOCA salt treatment was as described previously¹⁸.

Expression analysis

Total RNA was prepared from rat kidneys using tri-reagent. 20 µg of RNA was resolved on a 0.8% formamide gel, blotted to nylon membrane (Amersham) and probed with a partial rat ACE2 cDNA clone (9-1). The β-actin probe and multiple-tissue northern blots were purchased from Clontech. For western analysis, kidneys were homogenized in lysis buffer (50 mM Tris-HCl, pH 7.4, 20 mM EDTA, and 1% Triton-X100) supplemented with "Complete" protease inhibitor cocktail (Roche) and 1 mM Na₃VO₄. 100 µg of protein was resolved by SDS-polyacrylamide gel electrophoresis (PAGE) on 8% tris-glycine gels. ACE2 immuno-serum was obtained from rabbits immunized with a mouse-specific ACE2 peptide: DYEAEGADGYNYNRLIED. The serum was affinity-purified with the immunizing peptide using sulpho-link kit (Pierce). A commercially available β-actin antibody was used as loading control (Santa Cruz).

Generation of ACE2 mutant mice

A targeting vector (559 base pair short arm and 8.1 kilobase long arm) was constructed using the pKO Scrambler NTKV-1907 vector (Stratagene). A portion of the *ace2* genomic DNA containing nucleotides +1069 to +1299 was replaced with the neomycin resistance cassette in the anti-sense orientation. The targeting construct was electroporated into E14K embryonic stem cells, and screening for positive homologous recombinant embryonic stem cell clones was performed by Southern blotting of *Eco*RI-digested genomic DNA hybridized to 5' and 3' flanking probes. Two independent *ace*^{−/y} embryonic stem cell lines were injected into C57BL/6-derived blastocysts to generate chimaeric mice, which were backcrossed to C57BL/6 mice. Two embryonic stem cell lines gave independent germline transmission. Histology of all tissues, apoptosis assays, blood serology and kidney morphometries were as described³⁸. Complete ACE mutant mice have been previously described⁸ and were obtained from Jackson Laboratories. Mice were maintained at the animal facilities of the Ontario Cancer Institute in accordance with institutional guidelines.

Heart parameter measurements

For heart morphometry, hearts were perfused with 10% buffered formalin at 60 mm Hg

(1 mm Hg = 133.3 Pa) and subsequently embedded in paraffin. Myocardial interstitial fibrosis was determined by quantitative morphometry using the colour-subtractive computer-assisted image analysis using Image Processing Tool Kit version 2.5 coupled with Photoshop 6.0 software. Picro-Sirius-red (PSR)-stained sections were used to calculate interstitial fibrosis as the ratio of the areas with positive PSR staining compared to the entire visual field. Echocardiographic assessments were performed as described³⁹. Mice were anaesthetized with isoflurane/oxygen (1.25%/98.75%) and examined by transthoracic echocardiography using a Acuson Sequoia C256 equipped with a 15-MHz linear transducer. Haemodynamics measurements were performed as described⁴⁰. Tail-cuff blood pressure measurement were taken using a Visitech BP-2000 Blood Pressure Analysis System manufactured by Visitech Systems (Apex). For captopril treatment, drinking water was supplemented with 400 mg l^{−1} captopril (Sigma) for two weeks before blood pressure measurement.

Plasma and tissue angiotensins

For plasma measurements blood was collected into chilled Vacutainer blood collection tubes (Becton Dickinson) containing a mixture of peptidase inhibitors (25 mM ethylenediaminetetraacetic acid (EDTA), 0.44 mM *o*-phenanthroline (PHEN), 1 mM 4-chloromercuribenzoic acid (PCMB), 0.12 mM pepstatin A and 3 µM acetyl-His-Pro-Phe-Val-Statine-Leu-Phe, a renin inhibitor. Frozen hearts and kidneys (stored at −80 °C) were homogenized on ice in 80% ethanol/0.1 M HCl containing the peptidase inhibitors described above including phenylmethylsulphonyl fluoride (PMSF, 100 µM). A sample of the homogenate was taken to determine total protein content, using the Bradford protein assay with bovine serum albumin as a standard (BioRad Protein Assay Reagent, BioRad Laboratories). Peptide extraction was performed as previously described⁴¹. Radioimmunoassay analysis of angiotensin peptide content in the extracts from plasma, heart and kidney tissue was performed as previously described⁴². The limits of detection for the AngII, and AngI radio immunoassays were 0.5 fmol per tube, and 5 fmol per tube, respectively.

Drosophila studies

The P-element allele of *ACER*, P679 (ref. 31), was used to generate homozygous *ACER* mutant embryos. Whole embryo staining with Eve and Tin antibodies was as described³³.

Received 6 March; accepted 8 April 2002; doi:10.1038/nature00786.

1. Yusuf, S., Reddy, S., Ounpuu, S. & Anand, S. Global burden of cardiovascular diseases. Part I: General considerations, the epidemiologic transition, risk factors, and impact of urbanization. *Circulation* **104**, 2746–2753 (2001).
2. Carretero, O. A. & Oparil, S. Essential hypertension. Part I: Definition and etiology. *Circulation* **101**, 329–335 (2000).
3. Jacob, H. J. Physiological genetics: Application to hypertension research. *Clin. Exp. Pharm. Phys.* **26**, 530–535 (1999).
4. Rapp, J. P. Genetic analysis of inherited hypertension in the rat. *Physiol. Rev.* **80**, 135–172 (2000).
5. Stoll, M. *et al.* A genomic-systems biology map for cardiovascular function. *Science* **294**, 1723–1726 (2001).
6. Corvol, P. & Williams, T. A. in *Handbook of Proteolytic Enzymes* (eds Barrett, A. J., Rawlings, N. D. & Woessner, J. F.) 1066–1076 (Academic, London, 1998).
7. Skeggs, L. T., Dorer, F. E., Levine, M., Lentz, K. E. & Kahn, J. R. The biochemistry of the renin-angiotensin system. *Adv. Exp. Med. Biol.* **130**, 1–27 (1980).
8. Krege, J. H. *et al.* Male–female differences in fertility and blood pressure in ACE-deficient mice. *Nature* **375**, 146–148 (1995).
9. Esther, C. R. *et al.* Mice lacking angiotensin-converting enzyme have low blood pressure, renal pathology and reduced male fertility. *Lab. Invest.* **74**, 953–965 (1996).
10. Wuyts, B., Delanghe, J. & De Buyzere, M. Angiotensin I-converting enzyme insertion/deletion polymorphism: clinical implications. *Acta Clin. Belg.* **52**, 338–349 (1997).
11. Elkind, M. S. & Sacco, R. L. Stroke risk factors and stroke prevention. *Semin. Neurol.* **18**, 429–440 (1998).
12. Hollenberg, N. K. Angiotensin converting enzyme inhibition and the kidney. *Curr. Opin. Cardiol.* **3** (Suppl. 1), S19–S29 (1988).
13. Garg, R. & Yusuf, S. Overview of randomized trials of angiotensin-converting enzyme inhibitors on mortality and morbidity in patients with heart failure. *J. Am. Med. Assoc.* **273**, 1450–1456 (1995).
14. Tipnis, S. R. *et al.* A human homolog of angiotensin-converting enzyme. Cloning and functional expression as a captopril-insensitive carboxypeptidase. *J. Biol. Chem.* **275**, 33238–33243 (2000).
15. Donoghue, M. *et al.* A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1–9. *Circ. Res.* **87**, e1–e8 (2000).
16. Cornell, M. J. *et al.* Cloning and expression of an evolutionary conserved single-domain angiotensin converting enzyme from *Drosophila melanogaster*. *J. Biol. Chem.* **270**, 13613–13619 (1995).
17. Taylor, C. A., Coates, D. & Shirras, A. D. The *Acer* gene of *Drosophila* codes for an angiotensin-converting enzyme homologue. *Gene* **181**, 191–197 (1996).
18. Yagil, C. *et al.* Role of chromosome X in the Sabra rat model of salt-sensitive hypertension. *Hypertension* **33** Part II, 261–265 (1999).
19. Hilbert, P. *et al.* Chromosomal mapping of two genetic loci associated with blood-pressure regulation in hereditary hypertensive rats. *Nature* **353**, 521–529 (1991).
20. Kloting, I., Voigt, B. & Kovacs, P. Metabolic features of newly established congenic diabetes-prone BB.SHR rat strains. *Life Sci.* **62**, 973–979 (1998).
21. Koike, G. *et al.* Cloning, characterization, and genetic mapping of the rat type 2 angiotensin II receptor gene. *Hypertension* **26**, 998–1002 (1995).
22. Laragh, J. H. Renovascular hypertension: a paradigm for all hypertension. *J. Hypertens.* **4** (Suppl. 4), S79–S88 (1986).
23. Yagil, C. *et al.* Development, genotype and phenotype of a new colony of the Sabra hypertension prone (SBH/y) and resistant (SBN/y) rat model of salt sensitivity and resistance. *J. Hypertens.* **14**, 175–182 (1996).
24. Tanimoto, K. *et al.* Angiotensinogen-deficient mice with hypotension. *J. Biol. Chem.* **269**, 31334–31337 (1994).

25. Kloner, R. A., Bolli, R., Marban, E., Reinlib, L. & Braunwald, E. Medical and cellular implications of stunning, hibernation, and preconditioning: and NHLBI workshop. *Circulation* **97**, 1848–1867 (1998).
26. Murphy, A. M. *et al.* Transgenic mouse model of stunned myocardium. *Science* **287**, 488–491 (2000).
27. Heusch, G. Hibernating myocardium. *Physiol. Rev.* **78**, 1055–1085 (1998).
28. Sowter, H. M., Ratcliffe, P. J., Watson, P., Greenberg, A. H. & Harris, A. L. HIF-1-dependent regulation of hypoxic induction of the cell death factors BNIP3 and NIX in human tumors. *Cancer Res.* **61**, 6669–6673 (2001).
29. Kietzmann, T., Roth, U. & Jungermann, K. Induction of the plasminogen activator inhibitor-1 gene expression by mild hypoxia via a hypoxia response element binding the hypoxia-inducible factor-1 in rat hepatocytes. *Blood* **94**, 4177–4185 (1999).
30. Giordano, F. J. *et al.* A cardiac myocyte vascular endothelial growth factor paracrine pathway is required to maintain cardiac function. *Proc. Natl Acad. Sci. USA* **98**, 5780–5785 (2001).
31. Spradling, A. C. *et al.* The Berkeley *Drosophila* Genome Project gene disruption project: Single P-element insertions mutating 25% of vital *Drosophila* genes. *Genetics* **153**, 135–177 (1999).
32. Frasch, M., Hoey, T., Rushlow, C., Doyle, H. J. & Levine, M. Characterization and localization of the *even-skipped* protein of *Drosophila*. *EMBO J.* **6**, 749–759 (1987).
33. Azpiazu, N., Lawrence, P., Vincent, J.-P. & Frasch, M. Segmentation and specification of the *Drosophila* mesoderm. *Genes Dev.* **10**, 3183–3194 (1996).
34. Zhizhang, Y. & Frasch, M. Regulation and function of *tinman* during dorsal mesoderm induction and heart specification in *Drosophila*. *Dev. Gen.* **22**, 187–200 (1998).
35. Harvey, R. *NK-2* homeobox genes and heart development. *Dev. Biol.* **178**, 203–216 (1996).
36. Cai, H. & Harrison, D. G. Endothelial dysfunction in cardiovascular diseases: the role of oxidant stress. *Circ. Res.* **87**, 840–844 (2000).
37. Enseleit, F., Hurlimann, D. & Luscher, T. F. Vascular protective effects of angiotensin converting enzymes inhibitors and their relation to clinical events. *J. Cardiovasc. Pharmacol.* **37** (Suppl. 1), S21–S30 (2001).
38. Kong, Y. Y. *et al.* OPGL is a key regulator of osteoclastogenesis, lymphocyte development and lymph-node organogenesis. *Nature* **397**, 315–323 (1999).
39. Wickenden, A. D. *et al.* Targeted expression of a dominant-negative K(v)4.2 K(+) channel subunit in the mouse heart. *Circ. Res.* **85**, 1067–1076 (1999).
40. Zvaritch, E. *et al.* The transgenic expression of highly inhibitory monomeric forms of phospholamban in mouse heart impairs cardiac contractility. *J. Biol. Chem.* **275**, 14985–14991 (2000).
41. Allred, A. J., Chappell, M. C., Ferrario, C. M. & Diz, D. I. Differential actions of renal ischemic injury on the intrarenal angiotensin system. *Am. J. Physiol. Renal* **279**, F636–F645 (2000).
42. Chappell, M. C., Milsted, A., Diz, D. I., Brosnihan, K. B. & Ferrario, C. M. Evidence for an intrinsic angiotensin system in the canine pancreas. *J. Hypertens.* **9**, 751–759 (1991).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements

We thank D. Ganten for supplying us with tissue from SHRSP rats. Eve and Tin antibodies were a gift from M. Frasch. We acknowledge the Samuel Lunenfeld Research Institute's CMHD Mouse Physiology Facility for their technical screening services. This study was supported by Amgen and by grants from the Israel Science Foundation and the German-Israeli Foundation for Scientific Research and Development to C.Y. and Y.Y. J.M.P. holds a Canadian Research Chair in Cell Biology. M.A.C. is supported in part by a Canadian Institutes of Health Research fellowship.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.M.P. (e-mail: jpenning@uhnres.utoronto.ca).

A collimated jet of molecular gas from a star on the asymptotic giant branch

Hiroshi Imai*†‡, Kumiko Obara†§, Philip J. Diamond||, Toshinori Omodaka§ & Tetsuo Sasao*

* VERA Project Office, National Astronomical Observatory, Mitaka, Tokyo 181-8588, Japan

† Mizusawa Astrogeodynamics Observatory, Mizusawa, Iwate 023-0861, Japan

§ Department of Physics, Kagoshima University, Kagoshima 890-0065, Japan

|| Jodrell Bank Observatory, University of Manchester, Macclesfield, Cheshire SK11 9DL, UK

‡ Present address: Joint Institute for VLBI in Europe, Postbus 2, 7990 AA Dwingeloo, The Netherlands

Evolved stars of about one solar mass are in general spherically symmetric, yet the planetary nebulae that they produce in the next phase of their evolution tend not to exhibit such symmetry. Collimated 'jets' and outflows of material have been observed^{1–5} up to ~0.3 parsec from the central stars of planetary nebulae, and precession of those jets has been proposed¹ to explain the observed asymmetries. Moreover, it has recently been shown⁶ theoretically that magnetic fields could launch and collimate such jets. Here we report the detection of a collimated and precessing jet of molecular gas that is traced by water-vapour maser spots ~500 astronomical units (AU) from the star W43A in

Aquila. We conclude that the jet is formed in the immediate vicinity of the star, and infer that elongated planetary nebulae are formed by jets during the short period, of less than 1,000 years, when the star makes its transition through the proto-planetary nebula phase to become a planetary nebula.

W43A, lying at an estimated distance of 2.6 kpc from the Sun⁷, has been classified as an OH/infrared star, an evolved star on the asymptotic giant branch (AGB) of the Hertzsprung–Russell diagram, surrounded by a thick dust shell^{8,9}. Such stars are believed to be the immediate progenitors of proto-planetary nebulae, which rapidly evolve to become planetary nebulae^{9,10}. Maser (microwave amplification by stimulated emission of radiation) lines of three molecules (OH, H₂O and SiO) have been observed towards W43A^{7,11–14}, all with a central line-of-sight velocity of 34 km s^{–1} with respect to the local standard of rest (LSR). The 1,612-MHz OH masers exhibit a double-peaked spectral profile which varies in flux density with a period of about 400 days (ref. 8), caused by the periodic variation in infrared emission due to the stellar pulsation¹⁵. Although the true evolutionary status of W43A is still to be clarified, the characteristics of the OH masers, the detection of SiO masers¹⁴ and the lack of traditional tracers of young stellar objects⁷ suggest that W43A is an evolved star.

Previous observations of the H₂O and the OH masers have yielded a puzzling picture of the object. H₂O and OH masers are excited in gases with temperatures and hydrogen densities of $T \approx 400$ K, $n_{\text{H}_2} \approx 10^9$ cm^{–3} and $T \approx 50$ K, $n_{\text{H}_2} \approx 10^4$ cm^{–3}, respectively¹⁵, so it was natural that almost all known H₂O masers associated with evolved stars have been found closer to the star and have exhibited lower expansion velocities than OH masers^{16,17}. In

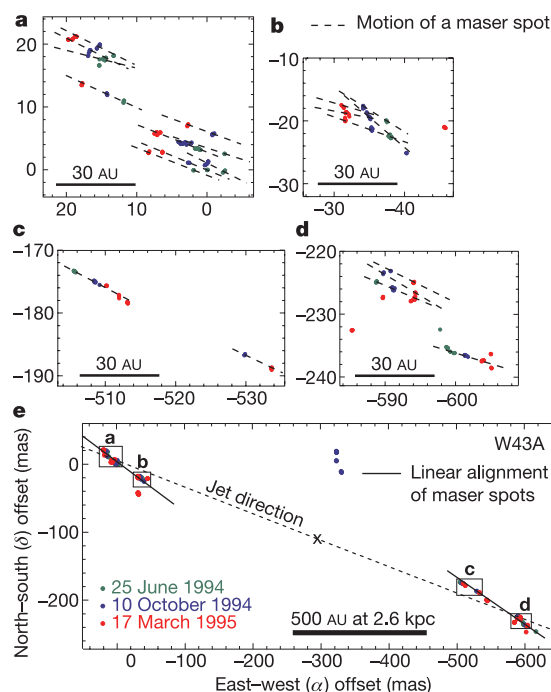


Figure 1 Time variation of the spatial distribution of H₂O masers toward W43A. H₂O maser emission (6₁₆–5₂₃ line at 22.235080 GHz) was observed with the VLBA with an angular resolution of 0.5 milliarcseconds (mas) in right ascension (α) and 1.0 mas in declination (δ). After signal correlation with the Socorro correlator, data with a velocity resolution of 0.21 km s^{–1} were obtained. Maser positions (green, blue and red dots) were measured in each epoch with respect to the brightest maser spot with typical uncertainties of 40 microarcseconds (μ as) in α and 60 μ as in δ . The map origin coincides with the position-reference spot in the first epoch, which was seen at all epochs at $\alpha(\text{J2000}) = 18^{\text{h}} 47^{\text{m}} 41.199^{\text{s}} \pm 0.002^{\text{s}}$, $\delta(\text{J2000}) = -01^{\circ} 45' 12.14'' \pm 0.614''$. At first, the maps from the three epochs were superposed so as to

locate the position-reference spot of the three maps at the same position in the superposed map. In close-ups of four sub-regions (labelled a–d), the majority of spots could be identified from epoch to epoch by checking the stability of their angular patterns and radial velocities: the variation being less than 1 km s^{–1} (ref. 13). In a–e, the three maps were again superposed by assuming that the map coordinate is moving with an average proper motion, $\mu_{\alpha} = -8.715$ mas per year, $\mu_{\delta} = -3.813$ mas per year, with respect to the position-reference spot. A dashed line in e and in Fig. 2 shows the direction of the jet, the linear fit of the alignment between the northeast and southwest maser clusters at position angle 69° east from north. A mark X in Figs 1–3 shows the dynamical centre of the jet (see also Fig. 3).

the case of W43A, however, although both the H₂O and OH masers show similar double peaked spectra centred around the same line-of-sight velocity, the H₂O maser peaks have a much larger velocity separation (about 180 km s⁻¹) than those of the OH maser (only about 16 km s⁻¹)^{11–13}. Low-resolution (0.1 arcsec) images of both species of maser¹² are clustered in two complexes: the approaching OH and receding H₂O lying to the northeast of the receding OH and approaching H₂O. The angular separation between the two H₂O complexes (around 0.5 arcsec or 1,300 AU) is twice as large as that of the OH and much larger than those typically observed in the circumstellar envelopes of many evolved stars (10–100 AU)^{16,18}.

Our new observations using the Very Long Baseline Array (VLBA) of the National Radio Astronomy Observatory revealed that the structure of the H₂O masers in W43A is highly collimated, both spatially and kinematically (Figs 1 and 2). Most of the maser spots are concentrated in the receding (northeast side) and approaching (southwest side) clusters. Both clusters have lengths of 250–350 AU but widths of only around 20 AU. The two clusters are separated by about 1,700 AU. The width of the aligned masers is only about 1/85 of the total length. The ratio is much smaller than any of those observed in “molecular outflows”^{3,19,20}. Measured proper motions with line-of-sight velocities of the masers indicate the presence of a bipolar jet with a three-dimensional space speed of 145 ± 20 km s⁻¹. Our analysis on the basis of diagonalization of the variance–covariance matrix of the velocity vectors²¹ shows that the vectors are concentrated towards a direction with a position angle of $62.7^\circ \pm 0.5^\circ$ east from north and an inclination of $-17.0^\circ \pm 3.0^\circ$ with respect to the sky plane. In this direction, the vectors are

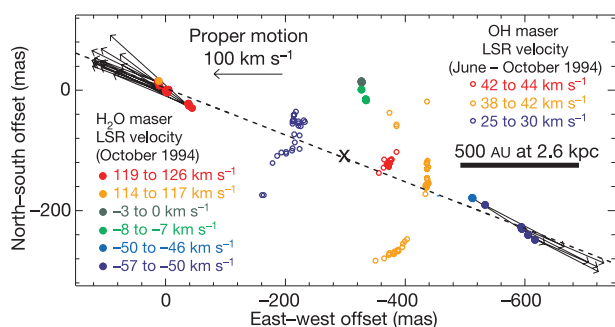


Figure 2 Spatio-kinematics of H₂O and OH masers in W43A. Assuming a point-symmetrical bipolar jet, the mean velocity vector of the receding motion was calculated to be 103.4 ± 11.3 km s⁻¹, 52.6 ± 16.1 km s⁻¹ and 86.5 ± 3.4 km s⁻¹ in α , δ and radial directions, respectively. The mean velocity vector of the approaching jet is opposite to that of the receding component. Out of observed OH maser emission (hyperfine transitions in the ground state at 1.612, 1.665, 1.667 and 1.720 GHz), the 1.612-, 1.665- and 1.667-GHz lines were detected and only the 1.612-GHz line, which was the strongest, could be imaged in the first and second epochs and shown here. After the correlation, OH data with a velocity resolution of 0.36 km s⁻¹ were obtained. The angular resolution was 9 mas and 21 mas and the relative positions of the masers were measured with a typical uncertainty of 1.2 mas and 2.2 mas in α and δ , respectively. The OH masers exhibit clear arcs and roughly fit a spherically expanding shell model with a radius of ~ 500 AU and an expansion velocity of ~ 9 km s⁻¹. These parameters are generally similar to, though slightly smaller than, those found in many OH/infrared stars^{7,8}. The dynamical age of the expanding shell was estimated to be around 2,600 years from the shell radius and the expansion velocity. Positions of the OH masers relative to the H₂O masers were estimated by assuming a common dynamical centre located at the middle point of the approaching ($V_{\text{LSR}} = 25\text{--}30$ km s⁻¹) and receding ($V_{\text{LSR}} = 42\text{--}44$ km s⁻¹) OH maser clusters with uncertainties of 30 mas in α and 50 mas in δ . This assumption should be reliable because the centre velocity of the spectra of both molecular species is the same to within 1 km s⁻¹, which corresponds to that of the common dynamical centre. We note that the off-jet H₂O masers shown in Fig. 1 appear to be located within the region occupied by the OH emission. LSR, local standard of rest.

9.5 ± 4.1 times larger than that on the plane perpendicular to the same direction. Moreover, two pairs of the sub-regions (a, d and b, c) are located almost point-symmetrically with respect to the estimated dynamical centre of the H₂O masers (X mark in Figs 1–3), which is the most likely origin of the jet. The extremely high collimation and point-symmetrical distributions of similar jets have been seen in optical observations of planetary nebulae and young stellar objects^{18,21,22}, but this is the first such molecular jet with these characteristics to be observed.

Theoretical models of H₂O masers suggest that maser excitation occurs in shocked regions created by collision between an outflow or a jet and the ambient gas cloud²³. H₂O masers have previously been observed on bow shocks around jets²⁴, on an arc of a spherically expanding flow²⁵ or in circumstellar envelopes where the masers show somewhat circularly symmetric distributions^{17,18,26}. Unlike all of these H₂O masers, most of the H₂O masers in W43A are likely to be excited in narrow locations close to the tip of the jet where material in the circumstellar envelope is being swept up and highly compressed to create the appropriate physical conditions for maser excitation.

Furthermore, the spatial alignment of H₂O masers in W43A does not constitute a simple straight line; the directions of the alignments in both the receding and approaching clusters are almost parallel to each other but shifted by about 10° from the directions connecting the two clusters, which in turn is almost parallel to the jet motions (Figs 1 and 2). We found that the observed angular pattern of the H₂O masers is well fitted by a precessing jet model (Fig. 3). Theoretical models suggest that a cosmic jet is created along the rotation axis of a gas disk around an object such as a protostar or a white dwarf so that the angular momentum of infalling material on the disk can be released from the compact volume^{19,27}. Adopting such models for W43A, the jet will exhibit wiggling owing to the precession of the disk.

Previous observations have inferred that stellar jets have already been formed approximately 1,000 years before an AGB star evolves into a post-AGB star and starts the photoionization of its circumstellar envelope⁵. H₂O masers in W43A are known to have existed for at least 25 years¹¹. However, the dynamical age of the jet in W43A is only around 28 years as estimated from the distances to the tips of the jet from the dynamical centre (about 900 AU) and the jet velocity

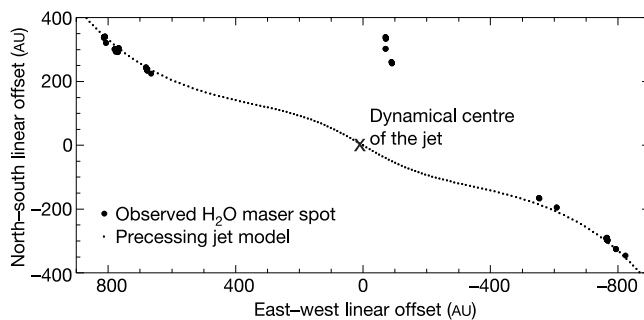


Figure 3 The angular distribution of the H₂O masers detected on 10 October 1994, fitted by the spiral pattern expected from a precessing jet model. The modelled jet has a constant velocity of 150 km s⁻¹ and a jet axis with an inclination of 39° with respect to the sky plane, position angle of 65° , and an axis precession with an angular amplitude of 5° and a period of 55 years. We assumed that only the direction of the bipolar ejection of material is varying continuously with time. The dynamical centre of the jet has a systemic radial velocity of 34 km s⁻¹ and a relative position of ($\Delta\alpha = -296$ mas, $\Delta\delta = -112$ mas) in Figs 1 and 2. Spiral patterns are also expected in the position-velocity (V_x , V_y , and V_z) plots, but no reliable fit to the observed velocity field was possible owing to its large uncertainty compared with the expected amplitude of variation of velocity vector (10 km s⁻¹).

(presumably constant at about 150 km s^{-1}). Combined with the presence of SiO masers in W43A¹⁴, which are detected in very few proto-planetary nebulae⁹, and with the OH masers such as those typically found in OH/infrared stars (Fig. 2), this estimate strongly suggests that we are observing a star in transition. It is likely that the “molecular jet” traced by H₂O masers was created in the OH/infrared star phase, before the transition to a proto-planetary nebula. Comparing the above timescales with the duration of significant mass-loss rate ($1,000\text{--}4,000$ years)²⁸ from an OH/infrared star (see also Fig. 2), a molecular jet should affect the development of nebula morphology from the beginning of its formation¹. However, the driving object and the formation mechanism of the jet are still unclear. W43A could be an AGB star able to create a collimated jet driven by the magnetic force owing to the dynamo action at the interface between the rapidly rotating core and the more slowly rotating envelope of the star⁶. Alternatively, W43A could be a binary system where the ejected material from a mass-losing star falls onto an accretion disk surrounding a companion that creates the jet^{29,30}. □

Received 19 October 2001; accepted 2 April 2002; doi:10.1038/nature00788.

- Sahai, R. & Trauger, J. T. Multipolar bubbles and jets in low-excitation planetary nebulae: toward a new understanding of the formation and shaping of planetary nebulae. *Astron. J.* **116**, 1357–1366 (1998).
- Sahai, R., te Lintel Hekkert, P., Morris, M., Zijlstra, A. & Likkell, L. The “Water-Fountain Nebula” IRAS 16342 – 3814: Hubble Space Telescope/Very Large Array study of a bipolar protoplanetary nebula. *Astrophys. J.* **514**, L115–L119 (1999).
- Alcolea, J., Bujarrabal, V., Sanchez, C. C., Neri, R. & Zweigle, J. The highly collimated bipolar outflow of OH 231.8 + 4.2. *Astron. Astrophys.* **373**, 932–949 (2001).
- Zijlstra, A. A. *et al.* Bipolar outflows in OH/IR stars. *Mon. Not. R. Astron. Soc.* **322**, 280–308 (2001).
- Miranda, L. F., Gómez, Y., Anglada, G. & Torrelles, J. M. Water-maser emission from a planetary nebula with a magnetized torus. *Nature* **414**, 284–286 (2001).
- Blackman, E. G., Frank, A., Markiel, J. A., Thomas, J. H. & Van Horn, H. M. Dynamos in asymptotic-giant-branch stars as the origin of magnetic fields shaping planetary nebulae. *Nature* **409**, 485–487 (2001).
- Diamond, P. J., Norris, R. P., Rowland, P. R., Booth, R. S. & Nyman, L.-A. The circumstellar envelopes around OH/IR stars. *Mon. Not. R. Astron. Soc.* **212**, 1–21 (1985).
- Herman, J. & Habing, H. J. Time variations and shell sizes of OH masers in late-type stars. *Astron. Astrophys.* **59**, 523–555 (1985).
- Nyman, L.-A., Hall, P. J. & Olofsson, H. SiO masers in OH/IR stars, proto-planetary and planetary nebulae. *Astron. Astrophys. Suppl. Ser.* **127**, 185–200 (1998).
- van der Veen, W. E. C. J. & Habing, H. J. The IRAS two-colour diagram as a tool for studying late stages of stellar evolution. *Astron. Astrophys.* **194**, 125–134 (1988).
- Genzel, R. & Downes, D. H₂O in the Galaxy: sites of newly formed OB stars. *Astron. Astrophys. Suppl.* **30**, 145–168 (1977).
- Diamond, P. J. & Nyman, L.-A. in *VLBI and its Impact on Astrophysics* (eds Moran, J. M. & Reid, M. J.) 249–250 (Astronomical Society of the Pacific, San Francisco, 1988).
- Likkell, L., Morris, M. & Maddalena, R. J. Evolved stars with high velocity H₂O maser features: bipolar outflows with velocity symmetry. *Astron. Astrophys.* **256**, 581–594 (1992).
- Nakashima, J. *Observational Study of O-rich Evolved Stars in the Galaxy*. PhD thesis, Graduate Univ. Advanced Studies (2002).
- Elitzur, M. *Astronomical Masers* (Kluwer Academic, Dordrecht, 1992).
- Cooke, B. & Elitzur, M. Water masers in late-type stars. *Astrophys. J.* **295**, 175–182 (1985).
- Chapman, J. M. & Cohen, R. J. MERLIN observations of the circumstellar envelope of VX Sagittarius. *Mon. Not. R. Astron. Soc.* **220**, 513–528 (1986).
- Lane, A. P. *et al.* H₂O masers in circumstellar envelopes. *Astrophys. J.* **323**, 756–765 (1987).
- Bachiller, R. Bipolar molecular outflows from young stars and protostars. *Annu. Rev. Astron. Astrophys.* **34**, 111–154 (1996).
- Chandler, F. & Guilleateau, S. The jet-driven molecular outflow of HH 211. *Astron. Astrophys.* **343**, 571–584 (1999).
- Bloemhof, E. E. Kinematics of the H₂O masers in W49N: Analysis of the velocity variance-covariance matrix. *Astrophys. J.* **533**, 893–896 (2000).
- Zinnecker, H., McCaughrean, M. J. & Rayner, J. T. A symmetrically pulsed jet of gas from an invisible protostar in Orion. *Nature* **394**, 862–865 (1998).
- Elitzur, M., Hollenbach, D. J. & McKee, C. F. Planar H₂O masers in star-forming regions. *Astrophys. J.* **394**, 221–227 (1992).
- Furuya, R. S. *et al.* A microjet: a protostar’s cry at birth. *Astrophys. J.* **542**, L135–L138 (2000).
- Torrelles, J. M. *et al.* Spherical episodic ejection of material from a young star. *Nature* **411**, 277–280 (2001).
- Marvel, K. B. The circumstellar environment of evolved stars as revealed by studies of circumstellar water masers. *Publ. Astron. Soc. Pacif.* **109**, 1286–1287 (1997).
- Uchida, Y. & Shibata, K. Magnetodynamical acceleration of CO and optical bipolar flows from the region of star formation. *Publ. Astron. Soc. Jpn* **37**, 515–535 (1985).
- Lewis, B. M. On the transience of high-latitude OH/IR stars. II. Thermal pulse link. *Astrophys. J.* **560**, 400–412 (2001).
- Morris, M. Mechanisms for mass loss from cool stars. *Publ. Astron. Soc. Pacif.* **99**, 1115–1122 (1987).
- Howard, B. E. in *Asymmetrical Planetary Nebulae II: From Origins to Microstructures*, ASP Conference Series (eds Kastner, H., Soker, N. & Rappaport, S.) Vol. 199 115–123 (Astronomical Society of the Pacific, San Francisco, 2000).

Acknowledgements

We thank J. Nakashima and S. Deguchi for providing the information concerning SiO maser emission and infrared images of W43A. NRAO is a facility of the National Science Foundation, operated under cooperative agreement by Associated Universities, Inc. H. I. was financially supported by the Research Fellowship of the Japan Society of the Promotion of Science for Young Scientist.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.I. (e-mail: imai@jive.nl).

Hidden orbital order in the heavy fermion metal URu₂Si₂

P. Chandra*, P. Coleman†, J. A. Mydosh‡ & V. Tripathi‡

* NEC, 4 Independence Way, Princeton, New Jersey 08540, USA

† Center for Materials Theory, Dept of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08855, USA

‡ Kamerlingh Onnes Laboratory, Leiden University, PO Box 9504, 2300 RA Leiden, The Netherlands

When matter is cooled from high temperatures, collective instabilities develop among its constituent particles that lead to new kinds of order¹. An anomaly in the specific heat is a classic signature of this phenomenon. Usually the associated order is easily identified, but sometimes its nature remains elusive. The heavy fermion metal URu₂Si₂ is one such example, where the order responsible for the sharp specific heat anomaly at $T_0 = 17 \text{ K}$ has remained unidentified despite more than seven-teen years of effort². In URu₂Si₂, the coexistence of large electron–electron repulsion and antiferromagnetic fluctuations leads to an almost incompressible heavy electron fluid, where anisotropically paired quasiparticle states are energetically favoured³. Here we develop a proposal for the nature of the hidden order in URu₂Si₂. We show that incommensurate orbital antiferromagnetism, associated with circulating currents between the uranium ions, can account for the local fields and entropy loss observed at the 17 K transition. We make detailed predictions for the outcome of neutron scattering measurements based on this proposal, so that it can be tested experimentally.

The intermetallic compound URu₂Si₂ contains a dense lattice of local moments where quantum fluctuations barely prevent spin ordering; the residual antiferromagnetic couplings between the strongly repulsive electrons are large and can drive new types of collective instabilities². At $T_0 = 17.5 \text{ K}$, URu₂Si₂ undergoes a second-order phase transition characterized by sharp discontinuities in bulk properties, including specific heat⁴, linear⁴ and non-linear^{5,6} susceptibilities, thermal expansion⁷ and resistivity⁸. The accompanying gap in the magnetic excitation spectrum⁹ suggests the formation of an itinerant spin density wave at this temperature; however, the size of the observed staggered moment¹⁰ ($m_0 = 0.03\mu_B$) cannot account for the entropy loss at this transition. The distinction between the primary hidden-order parameter and the secondary magnetic-order parameter is clarified by high-field measurements^{11,12} which indicate that the bulk anomalies survive up to an applied field strength of 40 tesla (T), whereas the magnetically ordered moment is destroyed by fields less than half this magnitude (15 T) (ref. 13). Furthermore, the size of

the small ordered moment grows linearly with pressure¹⁴, while the gap associated with the hidden order is relatively robust over this pressure range¹⁵.

There have been many theoretical proposals for the primary order parameter in URu₂Si₂ (ref. 16). Two recent nuclear magnetic resonance (NMR) studies have provided insights with crucial consequences for the nature of the hidden order. It has been widely assumed that the spin antiferromagnetism and the hidden order are coupled and homogeneous. However, a recent NMR study of URu₂Si₂ under pressure¹⁷ indicates that for $T < T_0$ there exist distinct antiferromagnetic and paramagnetic regions, implying that the magnetic and the hidden orders are phase-separated. This conclusion, supported by muon spin resonance (μ SR) data¹⁸, implies that the hidden-order phase contains no spin order. The observed growth of the staggered spin moment with applied pressure¹⁴ is then simply a volume-fraction effect which develops separately from the hidden order through a first-order transition¹⁹. At ambient pressure roughly a tenth or less of the system^{17,18} is magnetic, with $m_{\text{spin}} = 0.3\mu_B$. The main implication of these measurements for theory is that the magnetic and the hidden-order parameters are independent¹⁹.

In a parallel study, NMR measurements at ambient pressure²⁰ on URu₂Si₂ indicate that at $T \leq T_0$ the central (non-split) silicon NMR linewidth develops a field-independent, isotropic component whose temperature-dependent magnitude is proportional to that of the hidden-order parameter. These results imply an isotropic field distribution at the silicon sites whose root-mean-square value is proportional to the hidden order (ψ):

$$\langle B^\alpha(i)B^\beta(j) \rangle = A^2\psi^2\delta_{\alpha\beta} \quad (1)$$

and is about 10 gauss (G) at $T = 0$. This field magnitude is too small to be explained by the observed moment¹⁰ which induces a field $B_{\text{spin}} = (8/3)\pi Ma^{-3} = 100 \text{ G}$ where a is the U – U bond length ($a = 4 \times 10^{-8} \text{ cm}$). Furthermore, this moment is aligned along the c axis, and thus cannot account for the isotropic nature of the local field distribution detected by NMR. These measurements indicate that as the hidden order develops, an isotropically dis-

tributed static magnetic field develops at each silicon site. This is unambiguous evidence that the hidden-order parameter breaks time-reversal invariance.

Guided by these experiments, we now discuss our proposal for the nature of the hidden-order parameter. The magnetic fields at the silicon nuclei have two possible origins²¹: the conduction electron–spin interaction and the orbital shift that is due to current densities. In URu₂Si₂, the electron fluid exhibits a strong Ising anisotropy along the c axis, as measured by the Knight shift²⁰; hence, electron–spin coupling cannot be responsible for the observed isotropic fields. It would thus appear that these local fields are induced by currents that develop inside the crystal as the hidden order develops, and accordingly we attribute the observed isotropic linewidth to the orbital shift.

Here we propose that URu₂Si₂ becomes an incommensurate orbital antiferromagnet at $T = T_0$ with charge currents circulating between the uranium ions. Similar states have been studied extensively in the context of the two-dimensional Hubbard model^{22–26}, particularly in connection with staggered flux phases^{27,28}; more recently, commensurate current-density wave order has been proposed to explain the spin-gap phase in the underdoped copper oxide superconductors^{29,30}. The planar tetragonal structure of URu₂Si₂ lends itself naturally to an anisotropic charge instability of this type. Here we show that incommensurate orbital antiferromagnetic order in URu₂Si₂ can quantitatively account for the existing specific heat and NMR data for $T \leq T_0$. To test this, we make specific predictions for neutron scattering. Calculation of the structure factor requires knowledge of the fields throughout the full volume in real-space. The NMR only yields this information at discrete points, so we need some additional input to proceed. In the orbital antiferromagnet, the spatial dependence of the fields throughout the sample is determined by the requirement that the field distribution at the silicon sites is isotropic. Using this approach, we are able to link quantitatively the fields observed by NMR to the large specific heat anomaly that develops at T_0 . We also use the spatial field distribution associated with the incommensurate orbital antiferromagnet to predict the position, intensity and form-factor associated

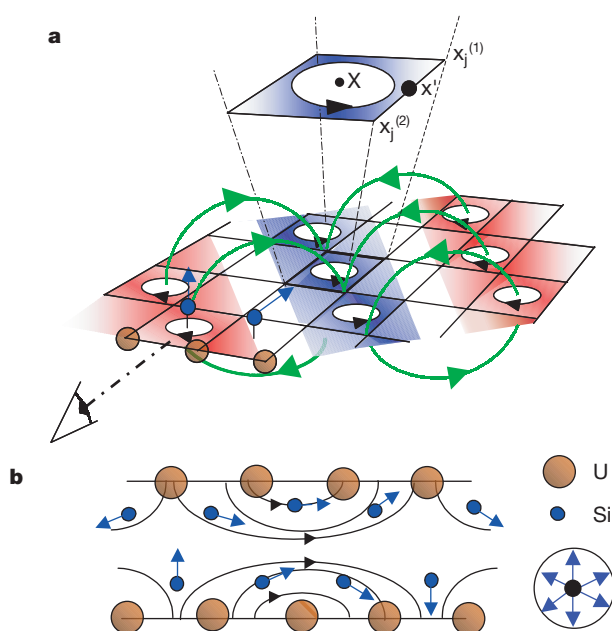


Figure 1 Magnetic field distribution associated with incommensurate orbital currents in the (0, 0, 1) plane. **a**, Schematic illustration of incommensurate orbital currents, showing resulting magnetic fields above and below the (0, 0, 1) plane. $x_j^{(1)}$ and $x_j^{(2)}$ are the

coordinates of the plaquette, and X' denotes the coordinate along a uranium–uranium bond. **b**, Side view showing how proposed field distribution is isotropic at silicon sites with a staggered current distribution between layers corresponding to $Q = (0.16, 0.16, 1)$.

with neutron scattering peaks on a surface of constant anisotropy in momentum space centred on the origin.

We begin by estimating the local fields at the silicon sites that are due to orbital currents circulating around the square uranium plaquettes in the a - b plane. On dimensional grounds, the current along the U - U bond is given by $I \approx e\Delta/\hbar$ where Δ is the gap associated with the formation of hidden order. Then the field induced at a height a above a plaquette is $B \approx (2/ac)(e\Delta/\hbar) = 11$ G, in good agreement with the local field strength detected in NMR measurements; here^{4,9-11} we have used $\Delta = 110$ K. The resulting orbital moment, $m_{\text{orb}} \approx 0.02\mu_B$, is comparable to the effective spin moment at ambient pressure ($m_0 = m_{\text{spin}} P_{\text{amb}} = 0.10 m_{\text{spin}} = 0.03\mu_B$). We emphasize that an orbital moment produces a local field an order of magnitude less than that associated with a spin moment of the same value; the low field strengths observed at the silicon sites are quantitatively consistent with our proposal that they originate from charge currents.

An orbital moment, $m_{\text{orb}} = 0.02\mu_B$, can also account for the sizable entropy loss at the transition. In a metal the change in entropy is given by $\Delta S = \Delta\gamma_n T_0$, where $\Delta\gamma_n$ is the change in the linear specific heat coefficient resulting from the gapping of the Fermi surface. $\Delta\gamma_n$ is inversely proportional to the Fermi energy ϵ_F of the gapped Fermi surface. Thus, in general, the change in entropy per unit cell is given by $\Delta S \approx k_B(k_B T_0/\epsilon_F)$. Exploiting the mean-field nature of this transition ($2\Delta \propto T_0$), we find that for the orbital antiferromagnet (OAFM):

$$\Delta S_{\text{OAFM}} \approx k_B \left(\frac{m_{\text{orb}}}{\mu_B} \right) = 0.02k_B \left(\frac{\mu_B^*}{\mu_B} \right) \quad (2)$$

where $\mu_B^* = (e/\hbar c)a^2\epsilon_F$ is the saturated orbital moment reached when the hidden-order gap Δ is approximately ϵ_F . The ratio $\mu_B/\mu_B^* = (a_0/a)^2(\epsilon_H/\epsilon_F)$, where a_0 is the Bohr radius and $\epsilon_H = e^2/(2a_0)$ is the energy of a hydrogen atom. For URu_2Si_2 , the very large size of $(\epsilon_H/\epsilon_F) \approx 10^3$ produced by the large mass renormalization of the heavy electrons actually offsets the small ratio $a_0/a \approx 10^{-1}$, so that $\mu_B/\mu_B^* \approx 10$, and $\Delta S_{\text{OAFM}} \approx 0.2k_B = 0.3k_B \ln 2$ is a number in good agreement with experiment⁴. The critical field for suppressing the associated thermodynamic anomalies is distinct from its spin counterpart; the ratio $H_c^{\text{orb}}/H_c^{\text{spin}} \approx \mu_B/\mu_B^* \approx 10$ is qualitatively consistent with the observed critical field of approximately 40 T associated with the destruction of the hidden order¹¹⁻¹³. From this simple discussion, we see that the sizable entropy loss associated with the development of the orbital antiferromagnetic state is a direct consequence of the strong renormalization of the electron mass in URu_2Si_2 ($m^*/m \propto \epsilon_H/\epsilon_F$). The absence of such a large effective mass ($m^*/m \approx 3$) in the copper oxides could explain why analogous thermodynamic anomalies are difficult to observe there.

Next we test whether the proposed orbital antiferromagnetism will yield the observed isotropic local fields. We allow the circulating current around a plaquette (see Fig. 1) centred at site X to develop a modulated magnetization $M(\mathbf{X}) = \psi e^{i\mathbf{Q} \cdot \mathbf{X}}$. The current along a bond is then the difference of the circulating currents along its adjacent plaquettes. The field at a silicon site can be computed using Ampere's law, where the relevant vector potential is:

$$\mathbf{A}(\mathbf{x}) = \frac{1}{c} \sum_j \int_{\mathbf{x}_j^{(1)}}^{\mathbf{x}_j^{(2)}} d\mathbf{x}' \frac{\mathbf{I}(\mathbf{x}_j)}{|\mathbf{x} - (\mathbf{x}_j + \mathbf{x}')|} \quad (3)$$

where $\mathbf{x}_j^{(1,2)}$ are the endpoints of the bond at site $\mathbf{x}_j$.

The silicon atoms in URu_2Si_2 are located at low-symmetry sites, so that the fields do not cancel; they reside above and below the centres of the uranium plaquettes. Microscopically, wavevector selection is most probably due to details of the Fermi surface; however, we can obtain a good idea of the likely modulation $\mathbf{Q}$ vector from the isotropic nature of the field distribution at the

silicon sites. For example, in the commensurate case $\mathbf{Q} = (1/2, 1/2, 1)$, the fields on the silicon sites are only along the c axis and thus the field distribution would be highly anisotropic. Consider a wavevector $(q, q, 1)$ (see Fig. 1). In this case, the currents are modulated within a plane with a wavelength $2\pi/q$, but staggered between planes. This then produces a circulating field distribution where the component of the field parallel to the $(0, 0, 1)$ planes becomes larger and larger as q is reduced. To obtain an isotropic field distribution, q needs to be reduced to a point where the horizontal and vertical components of the field are comparable. A detailed calculation based on the above model shows that the condition of perfect isotropy yields a circle of wavevectors (Fig. 2) centred around $\mathbf{Q} = (0, 0, 1)$ with a radius $q \approx 0.22$. Relaxation of this constraint results in an annulus of possible $\mathbf{Q}$ vectors, as shown in Fig. 2.

Our proposal of incommensurate current ordering in URu_2Si_2 can be tested by experiment. In particular we can Fourier-transform the real-space magnetic fields to calculate the neutron-scattering cross-section:

$$\frac{d\sigma}{d\Omega} = \left(\frac{g_N e}{8\pi\hbar c} \right)^2 |\mathbf{B}(\mathbf{q})|^2 = r_0^2 S(\mathbf{q}) \quad (4)$$

Here $|\mathbf{B}(\mathbf{q})|^2$ is the structure factor of the magnetic fields produced by the orbital currents and $S(\mathbf{q}) = |\mathbf{B}(\mathbf{q})|^2 / (4\pi\mu_B)^2$ is the structure factor of the orbital magnetic moments, measured in units of electron Bohr magnetons (μ_B). The factor $r_0 = g_N e^2 / 4m_e c^2$, m_e is the electron mass and g_N is the neutron gyromagnetic ratio. Using the vector potential from equation (3), we have calculated the magnetic field distribution for the incommensurate orbital antiferromagnet described above, and find that its Fourier transform is given by:

$$\mathbf{B}(\mathbf{q}) = \frac{4\pi}{c} N I a^2 \sum_{\mathbf{G}_{n_1, n_2, n_3}} \delta_{\mathbf{q}, \mathbf{Q} + \mathbf{G}} j_0 \left[\frac{q_x a}{2} \right] j_0 \left[\frac{q_y a}{2} \right] (1 + (-1)^{(n_1 + n_2 + n_3)}) \times \left(\left(\frac{q_y \hat{x} - q_x \hat{y}}{2q} \right) \times \hat{q} \right) \quad (5)$$

where $j_0(x) = (\sin x)/x$. From this expression, we can determine the intensities and the form factors associated with the diffraction. We find that there exists a set of dominant peaks associated with a constant anisotropy locus of wavevectors (Fig. 2) in the first Brillouin zone where $S(\mathbf{q}) = 0.18 (N I a^2 / c \mu_B)$, indicating that roughly a fifth of the total integrated weight of $S(\mathbf{q})$ lies here. Using the sum rule that relates the total integrated weight to the square of the moment, and the fact that at ambient pressure only a tenth of the sample is (spin) magnetic, we find that the scattering peaks due to orbital ordering in the first Brillouin zone should have 1/50 the intensity of the analogous spin magnetic peaks at ambient pressure. We have also calculated that these peaks will have a form factor (see Fig. 2) that scales with wavevector as q^{-4} , where this power-law decay indicates an extended scattering source.

We end with remarks about the microscopic formation of these charge currents. At low temperatures, heavy electron materials form almost incompressible Fermi liquids. The large Coulomb repulsion between the Landau quasiparticles strongly suppresses on-site charge fluctuations, demanding nodes in the particle-hole wavefunction. The resulting anisotropically paired states are also favoured by the residual antiferromagnetic interactions that persist in the heavy electron fluid. Indeed, we believe that the same d -wave pairing that drives the superconducting transition in URu_2Si_2 at $T = 1.5$ K also plays an important role in the formation of the orbital antiferromagnetism. We have found that the development of such anisotropic charge-density wave pairing occurs naturally in a simple model of URu_2Si_2 with nearest-neighbour antiferromagnetic

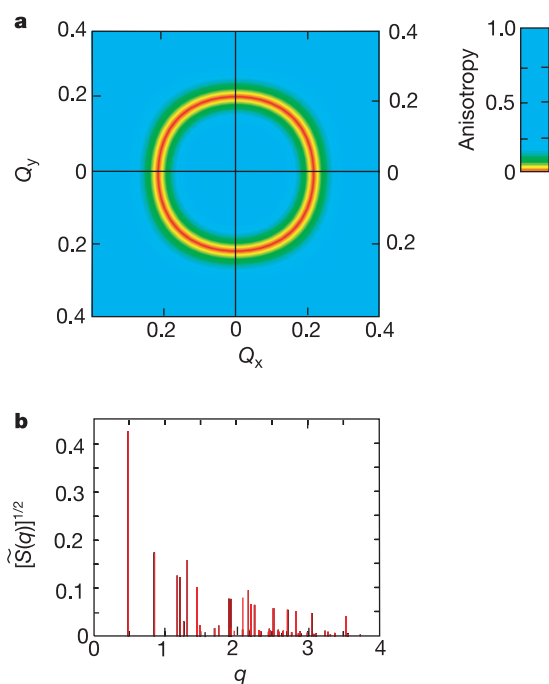


Figure 2 An isotropic magnetic field distribution at the silicon sites can be produced by many different wave vectors Q for orbital order, all of which give qualitatively similar neutron scattering patterns. **a**, Contour plot showing locus of constant anisotropy around a ring of radius $Q_{\perp} = 0.22$ in the vicinity of $Q = (0, 0, 1)$. **b**, Predicted elastic neutron scattering intensity, where $\tilde{S}(q)^{\frac{1}{2}} = S(q)^{\frac{1}{2}} / (\frac{N\mu_B^2}{C_{MB}}) \approx \sqrt{|B(q)|^2}$ gives a measure of the Fourier spectrum of magnetic fields $B(q)$, plotted for the representative case $Q = (0.16, 0.16, 1)$.

interactions:

$$H_I = \sum_{\mathbf{q}} J(\mathbf{q}) \mathbf{S}(\mathbf{q}) \cdot \mathbf{S}(-\mathbf{q}) \quad (6)$$

Here $\mathbf{S}(\mathbf{q})$ is the Fourier transform of the magnetization at wave-vector $\mathbf{q}$ and $J(\mathbf{q}) = 2J(\cos q_x + \cos q_y)$ for a nearest-neighbour interaction between adjacent U atoms in the basal plane. Expanding this interaction in terms of quasiparticle operators, we find that it can be rewritten as a sum of attractive interactions in four independent anisotropic charge-density wave channels:

$$H_I = -J \sum_{\Gamma=1,4, k_1, k_2, Q} (\gamma_{k_1}^{\Gamma})^* \gamma_{k_2}^{\Gamma} \rho_{k_1}(Q) \rho_{k_2}(-Q) \quad (7)$$

Here $\rho_k(Q) = \sum_{\sigma=\pm 1/2} c_{k-Q/2\sigma}^{\dagger} c_{k+Q/2\sigma}$ is the charge-density operator at wavevector Q , expressed in terms of creation and annihilation operators of the heavy quasiparticles. The form factors are $\gamma^{1,2}(\mathbf{k}) = \cos(k_x) \pm \cos(k_y)$, $\gamma^{3,4}(\mathbf{k}) = i(\sin(k_x) \pm \sin(k_y))$. Of these four possibilities, channels 1 and 3 do not have nodes, and are therefore suppressed by local Coulomb interactions. In the remaining two channels, only $\Gamma = 4$ breaks time reversal symmetry, giving rise to incommensurate orbital currents. One of the questions for further study concerns the excitations of this mean-field state. URu₂Si₂ is known to develop a dispersing singlet excitation¹⁰ at $T = T_0$. This mode was previously attributed to spin antiferromagnetism, now known to be absent from the hidden order phase¹⁷. We plan to study the possible identification of this propagating mode with the gapless phason associated with the uniform translation of an incommensurate orbital antiferromagnet.

Thus we have discussed the theoretical implications of two recent NMR experiments on the hidden order in URu₂Si₂. Pressure-dependent measurements indicate that it is completely independent of the spin magnetism in this material. We argue that the develop-

ment of isotropically distributed magnetic fields at the silicon sites at $T = T_0$ implies that the hidden-order parameter breaks time-reversal symmetry. The size and the isotropy of these observed local fields lead us to propose that URu₂Si₂ becomes an incommensurate orbital antiferromagnet at $T < T_0$. The heavy electron mass reduces the saturation value of the orbital moment, accounting for the sizable entropy loss at the transition and the scale of the associated critical field. We calculate the positions, intensities and form factor associated with the resulting neutron-scattering peaks so that this proposal can be tested by experiment. □

Received 30 January; accepted 29 April 2002; doi:10.1038/nature00795.

- Goodstein, D. L. *States of Matter* (Dover, New York, 1985).
- Buyers, W. J. L. Low moments in heavy-fermion systems. *Physica B* **223** & **224**, 9–14 (1996).
- Miyake, K., Schmitt-Rink, S. & Varma, C. M. Spin-fluctuation-mediated even-parity pairing in heavy fermion superconductors. *Phys. Rev. B* **34**, 6554–6556 (1986).
- Palstra, T. T. M. *et al.* Superconducting and magnetic transitions in the heavy fermion system URu₂Si₂. *Phys. Rev. Lett.* **55**, 2727–2730 (1985).
- Miyako, Y. *et al.* Magnetic properties of U(Ru_{1-x}Rh_x)Si₂ single crystals ($0 < x < 1$). *J. Appl. Phys.* **70**, 5791–5793 (1991).
- Ramirez, A. P. *et al.* Nonlinear susceptibility as a probe of tensor spin order in URu₂Si₂. *Phys. Rev. Lett.* **68**, 2680–2683 (1992).
- De Visser, A. *et al.* Thermal expansion and specific heat of monocrystalline URu₂Si₂. *Phys. Rev. B* **34**, 8168–8171 (1986).
- Palstra, T. T. M., Menovsky, A. A. & Mydosh, J. A. Anisotropic electrical resistivity of the magnetic heavy-fermion superconductor URu₂Si₂. *Phys. Rev. B* **33**, 6527–6530 (1986).
- Mason, T. E. & Buyers, W. J. L. Spin excitations and the electronic specific heat of URu₂Si₂. *Phys. Rev. B* **43**, 11471–11473 (1991).
- Broholm, C. *et al.* Magnetic excitations in the heavy-fermion superconductor URu₂Si₂. *Phys. Rev. B* **43**, 12809–12822 (1991).
- Mentink, S. A. M. *et al.* Gap formation and magnetic ordering in URu₂Si₂ probed by high-field magnetoresistance. *Phys. Rev. B* **53**, R6014–R6017 (1996).
- van Dijk, N. H. *et al.* Specific heat of heavy-fermion URu₂Si₂ in high magnetic fields. *Phys. Rev. B* **56**, 14493–14498 (1997).
- Mason, T. E. *et al.* Nontrivial magnetic order in URu₂Si₂? *J. Phys. Condens. Mat.* **7**, 5089–5096 (1995).
- Amitsuka, H. *et al.* Effect of pressure on tiny antiferromagnetic moment in the heavy-electron compound URu₂Si₂. *Phys. Rev. Lett.* **83**, 5114–5117 (1999).
- Fisher, R. A. *et al.* Specific heat of URu₂Si₂: Effect of pressure and magnetic field on the magnetic and superconducting transitions. *Physica B* **163**, 419–423 (1990).
- Shah, N., Chandra, P., Coleman, P. & Mydosh, J. A. Hidden order in URu₂Si₂. *Phys. Rev. B* **61**, 564–569 (2000).
- Matsuda, K. *et al.* Spatially inhomogeneous development of antiferromagnetism in URu₂Si₂: Evidence from ²⁹Si NMR under pressure. *Phys. Rev. Lett.* **87**, 087203–1–087203–4 (2001).
- Luke, G. M. *et al.* Muon spin relaxation in heavy fermion systems. *Hyper. Inter.* **85**, 397–409 (1994).
- Chandra, P., Coleman, P. & Mydosh, J. A. Pressure-induced magnetism and hidden order in URu₂Si₂. Preprint cond-mat/0110048 at (<http://xxx.lanl.gov>) (2001).
- Bernal, O. O. *et al.* ²⁹Si NMR and hidden order in URu₂Si₂. *Phys. Rev. Lett.* **87**, 153–156 (2001).
- Schlichter, C. P. *Principles of Magnetic Resonance* (Springer, Berlin, 1978).
- Halperin, B. I. & Rice, T. M. in *Solid State Physics* (eds Seitz, E., Turnbull, D. & Ehrenreich, H.) Vol. 21, 116–192 (Academic, New York, 1968).
- Affleck, I. & Marston, J. B. Large- n limit of the Heisenberg-Hubbard model: Implications for high T_c superconductors. *Phys. Rev. B* **37**, 3774–3777 (1988).
- Kotliar, G. Resonating valence bonds and D-wave superconductivity. *Phys. Rev. B* **37**, 3664–3666 (1988).
- Nersisyan, A. A. & Vachnadze, G. E. Low-temperature thermodynamics of two-dimensional orbital antiferromagnet. *J. Low-Temp. Phys.* **77**, 293–303 (1989).
- Schulz, H. J. Fermi surface instabilities of a generalized two-dimensional Hubbard model. *Phys. Rev. B* **39**, 2940–2943 (1989).
- Hsu, T. C., Marston, J. B. & Affleck, I. Two observable features of the staggered-flux phase at nonzero doping. *Phys. Rev. B* **43**, 2866–2877 (1991).
- Lee, P. A. Orbital currents in underdoped cuprates. Preprint cond-mat/0201052 at (<http://xxx.lanl.gov>) (2002).
- Chakravarty, S., Laughlin, R. B., Morr, D. K. & Nayak, C. Hidden order in the cuprates. *Phys. Rev. B* **63**, 094503–1–094503–10 (2001).
- Kee, H.-Y. & Kim, Y. B. Specific heat anomaly in the D-density wave state and emergence of incommensurate orbital antiferromagnetism. Preprint cond-mat/0111461 at (<http://xxx.lanl.gov>) (2001).

Acknowledgements

We acknowledge discussions with G. Aeppli, H. Amitsuka, O. Bernal, S. Chakravarty, L.P. Gorkov, G. Lonzarich, K. McEuen, D. McLaughlin, D. Morr and C. Nayak. P. C. and V. T. are supported by the National Science Foundation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.C. (e-mail: coleman@physics.rutgers.edu).

Ultrafast and direct imprint of nanostructures in silicon

Stephen Y. Chou*, Chris Keimel & Jian Gu

NanoStructure Laboratory, Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08544, USA

The fabrication of micrometre- and nanometre-scale devices in silicon typically involves lithography and etching. These processes are costly and tend to be either limited in their resolution or slow in their throughput¹. Recent work has demonstrated the possibility of patterning substrates on the nanometre scale by 'imprinting'^{2,3} or directed self-assembly⁴, although an etching step is still required to generate the final structures. We have devised and here demonstrate a rapid technique for patterning nanostructures in silicon that does not require etching. In our technique—which we call 'laser-assisted direct imprint' (LADI)—a single excimer laser pulse melts a thin surface layer of silicon, and a mould is embossed into the resulting liquid layer. A variety of structures with resolution better than 10 nm have been imprinted into silicon using LADI, and the embossing time is less than 250 ns. The high resolution and speed of LADI, which we attribute to molten silicon's low viscosity (one-third that of water), could open up a variety of applications and be extended to other materials and processing techniques.

In LADI, a single XeCl excimer laser pulse (308 nm wavelength and 20 ns pulse duration) passes through a quartz mould (which

does not absorb the laser energy because it has a bandgap larger than the photon energy) and melts a thin surface layer of the silicon substrate within picoseconds⁵. The molten silicon layer (which can be about 300 nm deep and remain molten for hundreds of nanoseconds) is then embossed by the quartz mould (Fig. 1). After the liquid silicon solidifies, the mould is separated from the imprinted silicon for the next imprint.

The experimental details are follows. First, the moulds, made of fused quartz 1 mm thick, have patterns on the quartz surface ranging from 10 nm to tens of micrometres in size that were fabricated by nanoimprint lithography and reactive ion etching. The moulds were diced into 1.5 mm × 1.5 mm pieces to fit within the excimer laser beam area of 2.5 mm × 2.5 mm, ensuring that all the silicon beneath the mould melts during LADI. The patterns on a LADI mould are of three types: (1) the 300 nm period grating of 140 nm linewidth and 110 nm depth; (2) the 10 nm wide and 15 nm deep lines, which were created as a result of the trenching effect in reactive ion etching during the mould fabrication; and (3) the rectangles of length and width in tens of micrometres and a depth of 110 nm.

Second, the pressure between the mould and silicon wafer was applied by sandwiching them between two large press plates. The silicon wafer was placed on the lower plate with the mould on top of the silicon wafer. The top plate, also made of fused quartz and hence transparent to the laser beam, was placed on top of the mould. Pressure was provided by screws between the two large plates. The exact pressure between the mould and the silicon wafer is unknown, but can be estimated from imprint time, imprint depth, and the mass of the mould, as discussed later. In this set-up the pressure is applied before the silicon is melted.

Third, the measured transmittance of the mould for the laser radiation is 93%, which indicates that the mould indeed does not absorb the laser energy, because the total reflection due to the

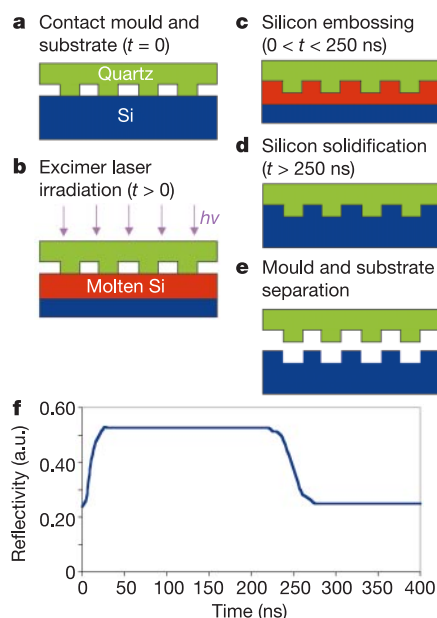


Figure 1 Schematic of laser-assisted direct imprint (LADI) of nanostructures in silicon. **a**, A quartz mould is brought into contact with the silicon substrate. A force presses the mould against the substrate. **b**, A single XeCl (308 nm wavelength) excimer laser pulse (20 ns pulse width) melts a thin surface layer of Si. **c**, The molten silicon is embossed while the silicon is in the liquid phase. **d**, The silicon rapidly solidifies. **e**, The mould and silicon substrate are separated, leaving a negative profile of the mould patterned in the silicon. **f**, The reflectivity of a HeNe laser beam from the silicon surface versus the time, when the silicon surface is irradiated by a single XeCl (308 nm) laser pulse with 1.6 mJ cm^{-2} fluence and 20 ns pulse duration. Molten Si, becoming a metal, gives a higher reflectivity. The measured reflectivity shows the silicon in liquid state for about 220 ns.

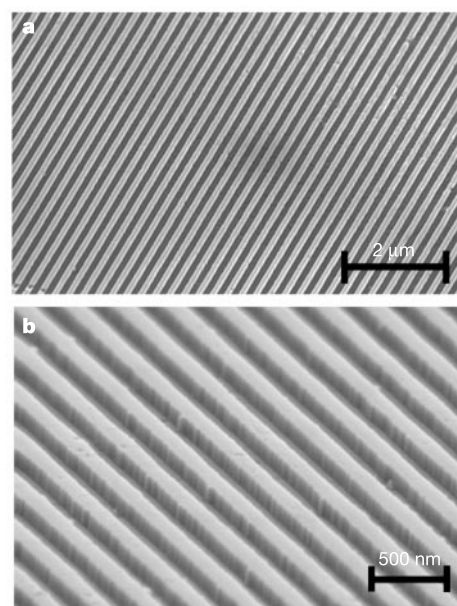


Figure 2 Scanning electron microscope (SEM) images. **a**, A uniform 300 nm period silicon grating patterned by LADI. The grating has 140 nm linewidth and is 110 nm deep. **b**, The mould after the two LADI processes showing no visible damage. During LADI the quartz mould does not melt, because quartz has a melting temperature about 300 °C higher than Si (quartz is used as the crucible material in Si-melting furnaces) and quartz conducts heat much less well—two orders of magnitude more poorly—than Si. Moreover, the quartz mould does not bind to Si during LADI.

difference in refractive index at the two mould–air interfaces should be around 7%.

Fourth, the melting of silicon surface can be monitored *in situ* during the LADI process, by measuring the time-resolved reflectivity of a HeNe laser beam (wavelength $\lambda = 633$ nm) from the silicon surface. When silicon melts, it changes from a semiconductor to a metal, hence its surface reflectivity to visible light increases by about a factor of two⁶. Figure 1f shows that under the laser radiation, the reflectivity of the Si surface increases rapidly in the first 25 ns, then saturates for 200 ns, and finally returns to the original solid silicon reflectivity in 50 ns. The silicon starts to melt immediately in less than picoseconds after the laser hit the surface⁷, so the reflectivity raise time should be the same as the laser pulse duration (20 ns), rather than 25 ns as measured. The discrepancy might be related to the RC (resistor–capacitor) time constant of the oscilloscope (100 MHz), which is 10 ns. The reflectivity measurements showed that the silicon surface was kept in liquid phase for about 220 ns. On

the basis of theoretical calculations⁷ and the melting depths experiments by other groups^{8,9}, the melting depth was estimated to be about 280 nm.

Using the time-resolved reflectivity measurements, we also investigated the effect of laser beam fluence on the silicon melt. We found that a fluence lower than 0.8 J cm^{-2} does not melt the silicon surface, but for a fluence higher than 2 J cm^{-2} , laser ablation of silicon will occur. We found that a laser pulse of 20 ns duration and 1.6 J cm^{-2} fluence melts a silicon surface sufficiently without ablation and has been used for all the results shown here.

The nanostructures directly imprinted in Si were characterized using scanning electron microscopy (SEM), atomic force microscopy (AFM) and optical microscopy. We found that the entire mould ($1.5 \text{ mm} \times 1.5 \text{ mm}$) was imprinted into the silicon wafer. Large-area uniform 300 nm period gratings in silicon by LADI have been achieved (Fig. 2a). The mould after two times of use does not show observable damage (Fig. 2b).

The cross-sectional views of the mould and the imprinted silicon features show that the imprinted Si structures are consistent with the mould (Fig. 3). One interesting feature in Fig. 3b is the ridge formed along the top corners on the imprinted grating lines. These ridges are about 10 nm wide and 15 nm high. The comparison with the mould indicates that these ridges come from the notches formed in the mould, which were caused by the trenching effect during reactive ion etching of the mould. The complete transfer of these 10-nm ridges in a LADI process suggests that the resolution of LADI is better than 10 nm.

Besides nanoscale features, LADI also demonstrated that isolated mesas and trenches of size of over tens of micrometres can be patterned, such as an isolated square mesa of $8 \mu\text{m}$ side and 110 nm height (Fig. 4). Successful imprinting of these large patterns indicates that the molten silicon can easily flow over tens of micrometres within nanoseconds.

The three intriguing characteristics of LADI—sub-10-nm resolution, sub-250-ns processing time, and excellent imprint of large

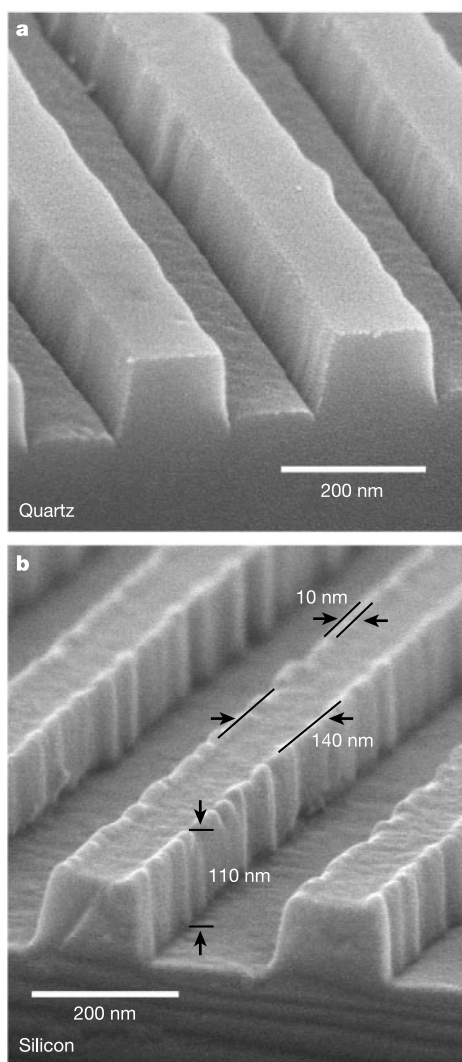


Figure 3 SEM image of the cross-section of samples patterned using LADI. **a**, A quartz mould. **b**, Imprinted patterns in silicon. The imprinted silicon grating is 140 nm wide, 110 nm deep and has a 300 nm period, an inverse of the mould. We note that the 10 nm wide and 15 nm tall silicon lines at each top corner of the silicon grating are the inverted replicas of the small notches on the mould (the notches were caused by the reactive ion etching trenching during mould fabrication). This indicates the sub-10-nm resolution of the LADI process. (These images are representative only. In fact, the Si structure in the image was probably not imprinted by the structure shown in the mould image.)

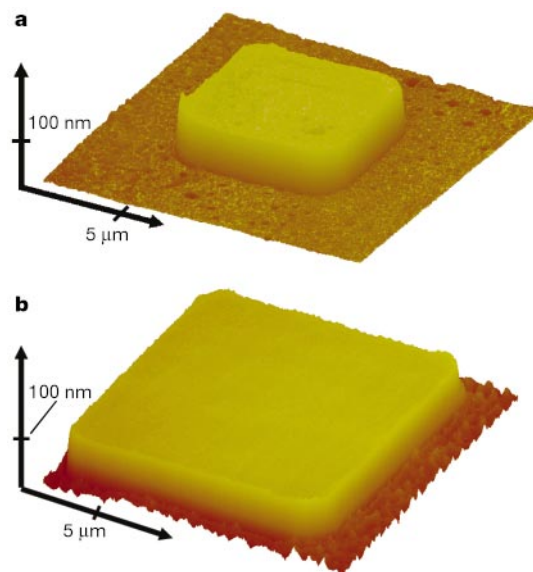


Figure 4 Atomic force micrographs (AFM) of isolated mesas patterned by LADI.

a, Crystalline silicon. **b**, Polysilicon. We note that the results of crystalline silicon and polysilicon are the same, except that the polysilicon has a slightly lower melting energy than crystalline silicon. The crystalline silicon mesa is $8 \mu\text{m}$ in length on each side and 110 nm tall. The large isolated pattern indicates that in LADI the liquid Si can flow over a large distance on the nanosecond scale. The polysilicon mesa is $17 \mu\text{m}$ in length on each side and 110 nm tall.

isolated patterns—are related to the fact that molten silicon has a viscosity of $0.003 \text{ cm}^2 \text{ s}^{-1}$, which is one-third that of water ($0.01 \text{ cm}^2 \text{ s}^{-1}$; ref. 10). This low viscosity enables the molten silicon to flow rapidly into all crevasses, filling them completely and conforming to the mould. Furthermore, silicon, like water, has a liquid phase density (2.52 g cm^{-3}) greater than its solid phase (2.32 g cm^{-3} ; ref. 11). The transformation from liquid to solid causes the silicon volume to expand about 3% in each direction. However, during LADI the mould was at a lower temperature than that of molten silicon, so the shrinking of silicon caused by the temperature drop can offset the expansion of silicon from liquid phase to the solid.

LADI can be applied to other materials. We have deposited, by chemical vapour deposition (CVD), a 230 nm polysilicon layer on top of a 200 nm thick silicon dioxide layer grown on a silicon substrate. Nanostructures have been patterned in the polysilicon layer by LADI (for example, Fig. 4b). The results are the same as that of crystalline silicon, except that the polysilicon has a slightly lower melting energy than crystalline silicon⁷. This indicates that LADI may become a good tool to directly pattern nanoscale gates for metal-oxide-semiconductor field-effect transistors (MOSFETs).

The velocity, acceleration, force, pressure, and Reynolds number involved in LADI of silicon can be estimated to provide further information about the process. In our experiments the imprint depth was 110 nm and the imprint time was around 250 ns. For simplicity, we assume that the mould travelled a distance of 100 nm in 200 ns; however, this distance may be slightly overestimated owing to elastic distortion of the mould under pressure and the upward flow of liquid silicon. This leads to an average imprint velocity of about 0.5 m s^{-1} and an average acceleration of $5 \times 10^6 \text{ m s}^{-2}$ —nearly a million times the gravitational acceleration. Because the mould has a weight of about 5.6 mg, the total force needed for the acceleration is about 28 N. For the given mould area (2.25 mm^2), the total pressure on the mould needed for the imprint is $1.7 \times 10^6 \text{ Pa}$ or about 17 atm. If we assume the liquid Si flow during LADI is one-dimensional and into a 140 nm opening at a speed of 0.5 m s^{-1} , then the Reynolds number is 0.23, which is quite small, indicating a laminar flow.

Finally, LADI can be extended to large areas, other materials, and other processes. The LADI area could be as large as a whole wafer (4 inch or 8 inch diameter), or a one-inch-square die (that die can be used to cover an entire wafer by step and repeat), provided that a uniform laser beam over a large area is available. Conventional nanoimprint has demonstrated excellent uniformity over a 4-inch wafer in a single step^{2,3}. LADI also could be used for other materials beyond crystalline silicon and polysilicon, such as Ge, III–V compound semiconductors and dielectrics (a different laser wavelength would be needed). LADI could help to crystallize polysilicon further. LADI might be well suited for three-dimensional patterning (for example, forming a lens on a Si or glass surface), which is challenging to achieve by conventional lithography and etching. LADI could offer a unique method to fill tiny holes in a dielectric (for example, silicon dioxide) with silicon, and a unique means of flattening the surface of a semiconductor deposited on a dielectric. Both are difficult issues in integrated circuit fabrication. Many applications of LADI are yet to be explored. □

Received 7 January; accepted 1 May 2002; doi:10.1038/nature00792.

1. Ito, T. & Okazaki, S. Pushing the limits of lithography. *Nature* **406**, 1027–1031 (2000).
2. Rogers, J. A. & Mirkin, C. Emerging methods for micro- and nanofabrication. *Mater. Res. Bull.* **26**, (2001).
3. Chou, S. Y., Krauss, P. R. & Renstrom, P. J. Imprint lithography with 25-nanometer resolution. *Science* **272**, 85–87 (1996).
4. Chou, S. Y. & Zhuang, L. Lithographically induced self-assembly of periodic polymer micropillar arrays. *J. Vac. Sci. Technol. B* **17**, 3197–3202 (1999).
5. Silvestrelli, P. L., Alavi, A., Parrinello, M. & Frenkel, D. Ab initio molecular dynamics simulation of laser melting of silicon. *Phys. Rev. Lett.* **77**, 3149–3152 (1996).
6. Ivlev, G. D. & Gatskevich, E. I. Liquid phase reflectivity under conditions of laser induced melting. *Semiconductors* **34**, 759–762 (2000).

7. Poute, J. M. & Mayer, J. *Laser Annealing of Semiconductors* (Academic, New York, 1982).
8. Carey, P. G. & Sigmon, T. W. In-situ doping of silicon using gas immersion laser doping (GILD). *Appl. Surf. Sci.* **43**, 325–332 (1989).
9. Weiner, K. H. & Sigmon, T. W. Thin-base bipolar transistor fabrication using gas immersion laser doping. *IEEE Electr. Device Lett.* **10**, 260–263 (1989).
10. Glazov, V. M., Chizhevskaya, S. N. & Glagoleva, N. N. *Liquid Semiconductors* (Plenum, New York, 1969).
11. Langen, M., Hibiya, T., Eguchi, M. & Egry, I. Measurement of the density and the thermal expansion coefficient of molten silicon using electromagnetic levitation. *J. Cryst. Growth* **186**, 550–556 (1998).

Acknowledgements

We thank Y. Zhan for his contributions in the initial phase of the work, Z. Suo for help with the discussion of the Reynolds number, and the US Defense Advanced Research Program Agency (DARPA), the Office of Naval Research (ONR) and the Army Research Office (ARO, through an equipment grant) for their partial financial support of the project.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.Y.C. (e-mail: chou@ee.princeton.edu).

Growth of early continental crust controlled by melting of amphibolite in subduction zones

Stephen Foley*, Massimo Tiepolo† & Riccardo Vannucci†

* Institut für Geologische Wissenschaften, Universität Greifswald, F.L. Jahnstrasse 17a, D-17487 Greifswald, Germany

† Dipartimento di Scienze della Terra, Università di Pavia and CNR- Istituto di Geoscienze e Georisorse (IGG) - Sezione di Pavia, Via Ferrata 1, I-27100 Pavia, Italy

It is thought that the first continental crust formed by melting of either eclogite or amphibolite, either at subduction zones¹ or on the underside of thick oceanic crust². However, the observed compositions of early crustal rocks and experimental studies have been unable to distinguish between these possibilities^{3–5}. Here we show a clear contrast in trace-element ratios of melts derived from amphibolites and those from eclogites. Partial melting of low-magnesium amphibolite can explain the low niobium/tantalum and high zirconium/samarium ratios in melts, as required for the early continental crust, whereas the melting of eclogite cannot. This indicates that the earliest continental crust formed by melting of amphibolites in subduction-zone environments and not by the melting of eclogite or magnesium-rich amphibolites in the lower part of thick oceanic crust. Moreover, the low niobium/tantalum ratio seen in subduction-zone igneous rocks of all ages is evidence that the melting of rutile-eclogite has never been a volumetrically important process.

The early continental crust is characterized by the tonalite-trondhjemite-granodiorite gneisses (TTG) of Archaean terrains³. Their compositions have been explained as the products of magmatic processes, namely melting of basaltic source rocks in the form of either eclogite or amphibolite^{1,4–5}. However, it is debated whether the melting process occurred principally in subduction zones¹ or on the underside of oceanic crust² that may have been much thicker in the Archaean owing to higher geothermal gradients.

TTG gneisses have low Nb/Ta and high Zr/Sm ratios relative to modern oceanic basalts and mantle rocks. They share these charac-

teristics with modern adakites (Fig. 1a), which are, on the basis of experimental petrology and major- and trace-element patterns, thought to originate by melting of the subducting basalt slab, and may thus be rare modern analogues of Archaean crust-building

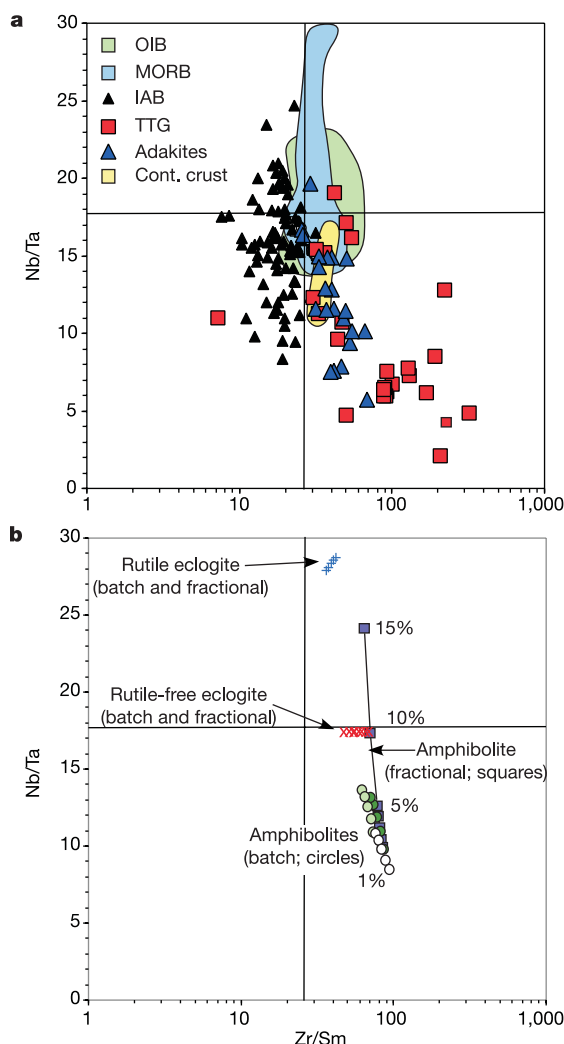


Figure 1 Nb/Ta ratios versus Zr/Sm ratios of natural rocks, compared to results of modelled melting of eclogites and amphibolites. **a**, The early continental crust represented by trondhjemite–tonalite–granodiorite gneisses (TTG) lie in the lower right quadrant with low Nb/Ta and high Zr/Sm, as do modern adakites. This trace-element signature differs from that of primitive mantle (at the intersection of the black lines), modern mid-ocean-ridge basalts (MORBs), ocean-island basalts (OIB) and island-arc basalts (IAB), all of which have lower Zr/Sm and/or higher Nb/Ta. The bulk continental crust also lies to the lower right, and may correspond to a TTG signature later diluted by crustal growth by other processes. **b**, Trace-element modelling of partial melting of eclogite and amphibolite using the new experimental partitioning data^{11–15} shows that melts of eclogite bearing rutile (blue plus signs) lie in the upper right quadrant, and eclogites without rutile (red crosses) have unfractionated Nb/Ta. Rocks in the lower right quadrant may originate by 1–15% batch melting of amphibolite. Melting models were based on high-pressure experiments in which mineral modes were measured^{4,5}; three different amphibolites were modelled, producing similar results: (1) 0.8 GPa (ref. 4) (white circles) with 40% amphibole (Am), 45% plagioclase (Plg) and 15% orthopyroxene; (2) 1.0 GPa (ref. 5) (light green circles) with 32% Am, 15% Plg, 28% garnet (Ga) and 25% clinopyroxene (Cpx); (3) 1.6 GPa (ref. 4) (dark green circles) with 17% Am, 15% Plg, 15% Ga and 52% Cpx. Purple squares are for pure fractional melting at 1.0 GPa, showing that only very low degree fractional melts could carry this trace-element signature. Partition coefficients used are for Plg with An_{45} and Am with $Mg\# = 45$. The starting composition used for modelling was an average Archaean basalt from Sula Mountain³⁰.

magmatism⁶. Subducting basalt would be much more likely to melt during Archaean times because of the higher geothermal gradient and probably lower average age of oceanic crust being subducted⁷. Any crust produced by this process would, therefore, have been much more voluminous than adakites are today. The alternative model of melting of lower levels of thickened oceanic crust has no direct modern analogue: thick oceanic plateaux do occur and serve as a model for the Archaean oceanic crust⁸, but the lower thermal gradients today prevent them from melting. Nevertheless, there are differences between these two models for the origin of TTG gneisses: subducting crust could conceivably melt as amphibolite or eclogite, but its average composition would generally be that of fractionated basalt; more than 90% of mid-ocean-ridge basalts (MORBs) have Mg number ($Mg\# = 100 \text{ Mg}/(\text{Mg} + \text{Fe})$) between 50 and 65. The underside of thick oceanic crust would have higher $Mg\#$ because it consists of a high proportion of cumulates⁸. Furthermore, the thick crust would be formed by high-degree melts with very low water contents⁹, and so the lower crust would consist largely of granulite and eclogite, not amphibolite. Hydrothermal alteration at ridges is

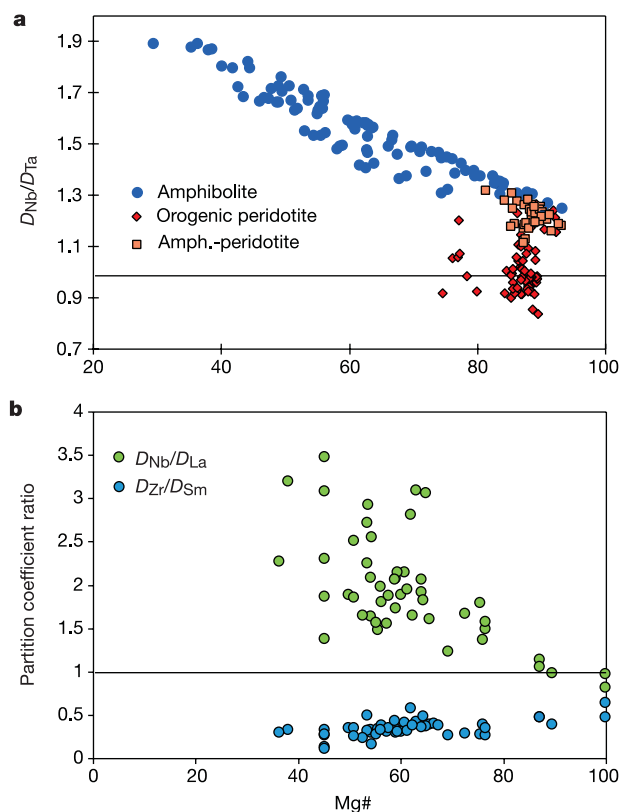


Figure 2 Trace-element ratios in amphiboles as a function of $Mg\#$ ($= 100 \text{ Mg}/(\text{Mg} + \text{Fe})$). **a**, D_{Nb}/D_{Ta} ratios are calculated using the equation given in the text together with the $Mg\#$ and Ti contents from microprobe analyses of natural and experimental amphiboles. Plotted as a function of $Mg\#$, we can see that only amphiboles with $Mg\#$ below about 70 can fractionate Nb from Ta significantly, producing the required low Nb/Ta in coexisting melts. Amphiboles with $Mg\#$ 80–90, modelled here by amphiboles in peridotites (natural and experimental), cannot fractionate Nb from Ta. This would apply to high- $Mg\#$ amphibolites that may result from metamorphism of mafic to ultramafic cumulates in the lower oceanic crust. **b**, Partition coefficient ratios from our experiments^{11–13} show that the lower- $Mg\#$ amphiboles also fractionate Nb from La (green circles) and Zr from Sm (blue circles), whereas amphiboles with $Mg\# > 80$ cannot. Melts derived by melting of amphibolites with low $Mg\#$ will therefore have a more significant trough in the incompatible element pattern at Nb and Ta. The values plotted in **b** are from the same experiments used to calibrate the predictive expression for D_{Nb}/D_{Ta} applied in **a**.

not likely to reach deeper than the uppermost 3 or 4 km (ref. 10).

We now contrast the behaviour of trace elements during melting of amphibolite and eclogite on the basis of recent experimental trace-element partitioning studies. Current models for the genesis of TTG, the continental crust and adakites require garnet in the residue to explain low concentrations of heavy rare-earth elements (HREE), for which both garnet amphibolite and eclogite melting provide reasonable solutions^{1,4}. The behaviour of the high-field-strength elements (HFSE) Nb, Ta, Zr and Hf are now sufficiently well known to assess critical ratios involving these elements. The amphibole partitioning data are taken from our detailed experimental studies at 1.4 GPa pressure, in which trace-element determinations in mineral-glass pairs in experimental run products by ion microprobe were combined with crystal structure refinements for all experimentally synthesized amphiboles, so that the site preferences and incorporation mechanisms of trace elements are well-defined^{11–13}. This is particularly important for understanding the partitioning behaviour of the HFSE, as it can be shown that amphibole decouples Nb and Ta from Zr and Hf. Nb and Ta are located on the M1 site, charge-balancing dehydrogenation of the amphiboles (exchange of O^{2-} for OH^-), whereas Zr (and Hf) are not related to dehydrogenation and occur on the M2 site^{12,13}. The partitioning behaviour for the eclogite minerals garnet, clinopyroxene and rutile is taken from experiments on tonalite^{14,15}, and those for the other amphibolite minerals plagioclase and orthopyroxene from previous studies^{16–20}. Almost all partition coefficients (D) used are obtained using *in situ* analyses of minerals and glass by either ion microprobe or laser ablation inductively coupled plasma mass spectrometry (ICP-MS).

An important result of the amphibole study is that the ionic radii of Nb and Ta are not identical as almost universally assumed, but that Ta is about 0.015 Å smaller¹². This means that in any mineral in which these two elements occupy a site smaller than Ta, low D_{Nb}/D_{Ta} (<1) will occur in that mineral: this is confirmed for rutile by experimental measurements where both partition coefficients have been determined^{21,22}. In contrast, the effective size of the M1 site in amphibole straddles the ionic radii of Nb and Ta, with the result that D_{Nb}/D_{Ta} is higher than 1 for low-Mg# amphiboles¹². Thus, where rutile dominates the Nb-Ta budget, as for rutile-bearing eclogites, high Nb/Ta is to be expected in melts.

In Fig. 1b, the Nb/Ta and Zr/Sm ratios of melts produced from amphibolites and eclogites are compared for the extreme cases of batch and pure fractional melting. Results are shown for eclogites with and without 0.5% rutile: the latter would be relevant only for exceptionally depleted basalt compositions, as recent experiments have shown that rutile saturation should be achieved during partial melting²³. This is consistent with the almost ubiquitous occurrence of rutile in Archaean eclogite xenoliths^{24,25}. For the amphibolites, measured modal mineralogies from suprasolidus experiments on melting of basalt at 0.8, 1.0 and 1.6 GPa were used (Fig. 1b)^{4,5}. Amphibole is a residual mineral in all of these. The 0.8-GPa assemblage is garnet-free and orthopyroxene-bearing, whereas the higher-pressure assemblages contain garnet and clinopyroxene in addition to amphibole and plagioclase. The vertical array of symbols (Fig. 1b) shows that all model melts can explain high Zr/Sm relative to mantle values, but that only amphibolite models can explain low Nb/Ta. The ratios for melts of eclogite, especially with rutile, do not fit that of TTG, the bulk continental crust, adakites or most modern island-arc basalts. Whereas melting of rutile-bearing eclogite can explain low HREE concentrations and high Zr/Sm, it fails the critical Nb/Ta test. Batch melts (1–15%) of all three amphibolites lie in the lower right quadrant, as do pure fractional melts at degrees of melting below 10% (all melts to 15% in the case the 0.8-GPa amphibolite). Batch melting may be the more realistic model given the relatively high SiO_2 content of the melts.

During melting of amphibolites, Nb/Ta in the melt is controlled by the Mg# and titanium concentration of the amphiboles, and can be predicted from microprobe analyses using the following expression: $\ln(A^{Amph/L} D_{Nb/Ta}) = 2.45 - 1.26Mg - 0.84Ti$ ($r^2 = 0.84$; $A^{Amph/L} D_{Nb/Ta}$ is the partition coefficient ratio for Nb and Ta between amphibole and melt; average difference between calculated and observed ratios was 7%, maximum 20%; ref. 12). Application of this equation to a compilation of amphibole compositions from amphibolites and peridotites shows that only amphiboles with Mg# less than about 70 can cause low Nb/Ta in coexisting melts (Fig. 2a), and that the fractionation increases with decreasing Mg# of the amphibole. These low-Mg# amphiboles are also able to cause the greatest fractionation between Nb/La and Zr/Sm (Fig. 2b). In contrast, amphiboles with Mg# > 80 cannot impart low Nb/Ta to coexisting melts, nor can they cause low Nb/La. We conclude that the trace-

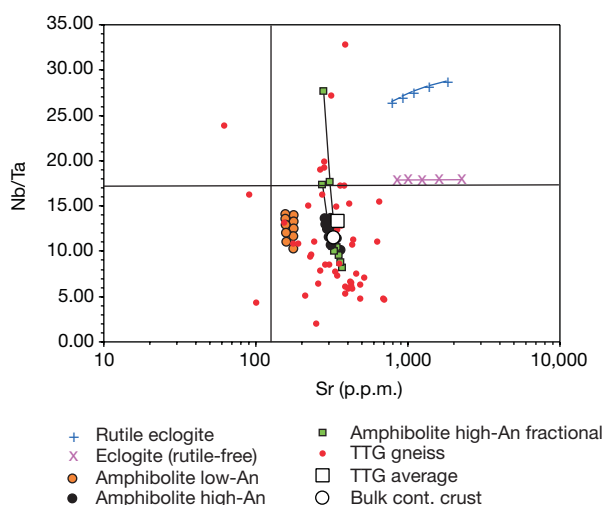


Figure 3 Modelling of Sr concentrations in melts of amphibolite and eclogite. The high calculated Sr concentrations in melts of eclogite, with (blue plus signs) or without rutile (pink crosses), further supports the conclusion that TTG (shown as an average and as individual rocks) cannot have originated by melting of eclogite. Melts from amphibolite fit better, whereby the Sr concentration depends on D_{Sr} , which itself is inversely proportional to the anorthite content of plagioclase¹⁹. The 'high-An' amphibolites which produce the

best fit may result from high water contents during melting²⁷. The starting composition used for modelling was the same average Archaean basalt as in Fig. 1b with 157 p.p.m. Sr (ref. 30). Use of a lower Sr concentration such as in modern MORB (113 p.p.m.) would not change the conclusions appreciably; the bulk partition coefficient for Sr is the controlling factor. The intersection of the black lines indicates primitive mantle values for reference.

element characteristics of TTG, including low Nb/Ta and Nb/La and high Zr/Sm, can be explained by melting of amphibolite in the subducting slab where amphiboles have relatively low Mg#, but not by melting of Mg-rich amphibolites or eclogites in the lower parts of thick oceanic crust. The decoupling of Nb and Ta from Zr and the fractionation of Nb from Ta in low-Mg# amphiboles is the best explanation for the production of melts in the lower right quadrant of Fig. 1; no other mantle phase with this capability has been reported.

Melting of rutile eclogite cannot be (or have been) volumetrically important in the production of igneous rocks at convergent margins, apart, perhaps, from the production of relatively rare alkaline volcanics with high Nb/Ta²⁶. Many eclogite xenoliths are thought to represent former oceanic crust that melted during subduction^{24,25}, and many have high Nb/Ta rutiles²⁵. Our results show that those with high Nb/Ta cannot have formed with rutile coexisting with melt, but must have melted as rutile-free amphibolites, followed by solid-state transformation of amphibolite to rutile eclogite on further subduction. This must also apply to the supposed lost eclogite reservoir with high Nb/Ta that may be located deep in the mantle²⁵.

Although melting of garnet amphibolite with amphibole Mg# of 40–50 and plagioclase with 40–50% anorthite end member (An_{40–50}) can explain the low HREE, high Zr/Sm and low Nb/Ta of the early continental crust, minor refinements to the model may be necessary for some trace elements, notably Sr, U and Th. The predicted melts from all amphibolite melting models shown in Fig. 1b have U and Th contents that are higher than TTG and the bulk continental crust, indicating that D_U and D_{Th} are too low in the models by a factor of about 3. This may be due to underestimation of the modal proportion of amphibole and/or plagioclase (see Fig. 1b), or to incorrect estimation of D_U and D_{Th} for plagioclase²⁰. The problem for the more abundant trace element strontium is more acute (Fig. 3). D_{Sr} is inversely proportional to the anorthite content of plagioclase, as parameterized by Blundy and Wood¹⁹. The 'low-An' circles in Fig. 3 are for the 1.0- and 1.6-GPa amphibolites from Fig. 1b with $D_{Sr} = 5.25$, corresponding to plagioclase with An₄₅. Using a lower D_{Sr} of 2, corresponding to a more Ca-rich plagioclase, excellent agreement results for Sr concentrations in average TTG and bulk continental crust in a model otherwise identical to the low-An model. This correction is not arbitrary, as it is notoriously difficult to attain equilibrium in experiments on plagioclase, and it has been shown that the presence of water results in a considerable increase in the anorthite content of plagioclase²⁷. The partitioning data used for plagioclase^{18,19} are based on experiments without water. The calculations for melting of eclogite in Fig. 3 confirm that the early continental crust is not related to melting of eclogite.

These results also eliminate the melting of amphibole peridotite²⁸ as a possible cause of the low Nb/Ta in modern island-arc volcanics (Fig. 1a), because the high Mg# of amphiboles from mantle wedge assemblages above the subducting slab prevents fractionation of Nb from Ta (Fig. 2a). A recent study of fluids released by antigorite breakdown in deeply subducted mantle shows that they do not display relative enrichment in large ion lithophile elements compared to HFSE, thus strongly supporting the hypothesis that HFSE are soluble in natural subduction fluids²⁹. Amphibole may nevertheless cause the low Nb/La and low Nb/Ta signatures of arc magmas: low-Mg#, low-Ti amphiboles may crystallize from slab-derived, silica-rich aqueous fluids at high fluid/rock ratios, causing Nb depletion relative to La and Ta in residual fluids. These move further and progressively equilibrate with peridotite mantle at lower fluid/rock ratios, while migrating to the regions where large-scale melting of the wedge occurs. During melting of amphibole peridotite, amphibole is no longer able to fractionate Nb from Ta or La to an appreciable extent. □

Received 4 October 2001; accepted 16 May 2002; doi:10.1038/nature00799.

- Drummond, M. S. & Defant, M. J. A model for trondhjemite-tonalite-dacite genesis and crustal growth via slab melting: Archean to modern comparisons. *J. Geophys. Res.* **95**, 21503–21521 (1990).
- Kröner, A. Evolution of the Archean continental crust. *Annu. Rev. Earth Planet. Sci.* **13**, 49–74 (1985).
- Martin, H. in *Archean Crustal Evolution* (ed. Condie, K. C.) 205–259 (Elsevier, Amsterdam, 1994).
- Rapp, R. P., Watson, E. B. & Miller, C. F. Partial melting of amphibolite/eclogite and the origin of Archean trondhjemites and tonalites. *Precamb. Res.* **51**, 1–25 (1991).
- Wolf, M. B. & Wyllie, P. J. Dehydration-melting of amphibolite at 10 kbar: the effects of temperature and time. *Contrib. Mineral. Petrol.* **115**, 369–383 (1994).
- Martin, H. Adakitic magmas: modern analogues of Archean granitoids. *Lithos* **46**, 411–429 (1998).
- Karsten, J. L., Klein, E. M. & Sherman, S. B. Subduction zone geochemical characteristics in ocean ridge basalts from the southern Chile ridge: implications of modern ridge subduction systems for the Archean. *Lithos* **37**, 143–161 (1996).
- Saunders, A. D., Tarney, J., Kerr, A. C. & Kent, R. W. The formation and fate of large oceanic igneous provinces. *Lithos* **37**, 81–95 (1996).
- Arndt, N. T. *et al.* Were komatiites wet? *Geology* **26**, 739–742 (1998).
- Bickle, M. J., Nisbet, E. G. & Martin, A. Archean greenstone belts are not oceanic crust. *J. Geol.* **102**, 121–138 (1994).
- Bottazzi, P. *et al.* Distinct site preferences for heavy and light rare-earth elements and the prediction of $^{147}\text{Sm}/^{149}\text{Sm}$. *Contrib. Mineral. Petrol.* **137**, 36–45 (1999).
- Tiepolo, M. *et al.* Nb and Ta incorporation and fractionation in titanite paragonite and kaersutite: crystal-chemical constraints and implications for natural systems. *Earth Planet. Sci. Lett.* **176**, 185–201 (2000).
- Tiepolo, M., Bottazzi, P., Foley, S. F., Overt, R. & Zanetti, A. Fractionation of Nb and Ta from Zr and Hf at mantle depths: the role of titanite-paragonite and kaersutite. *J. Petrol.* **42**, 221–232 (2001).
- Barth, M. G., Foley, S. F. & Horn, I. Partial melting in Archean subduction zones: constraints from experimentally determined trace-element partition coefficients between eclogitic minerals and tonalitic melts under upper mantle conditions. *Precamb. Res.* **113**, 323–340 (2002).
- Foley, S. F., Barth, M. G. & Jenner, G. A. Rutile/melt partition coefficients for trace elements and an assessment of the influence of rutile on the trace element characteristics of subduction zone magmas. *Geochim. Cosmochim. Acta* **64**, 933–938 (2000).
- Kennedy, A. K., Lofgren, G. E. & Wasserburg, G. J. An experimental study of trace element partitioning between olivine, orthopyroxene and melt in chondrules: equilibrium values and kinetic effects. *Earth Planet. Sci. Lett.* **115**, 177–195 (1993).
- Eggins, S. M., Rudnick, R. L. & McDonough, W. F. The composition of peridotites and their minerals: a laser-ablation ICP-MS study. *Earth Planet. Sci. Lett.* **154**, 53–71 (1998).
- Bindemann, I. N., Davis, A. M. & Drake, M. J. Ion microprobe study of plagioclase-basalt partition experiments at natural concentration levels of trace elements. *Geochim. Cosmochim. Acta* **62**, 1175–1193 (1998).
- Blundy, J. D. & Wood, B. J. Crystal-chemical controls on the partitioning of Sr and Ba between plagioclase feldspar, silicate melts, and hydrothermal solutions. *Geochim. Cosmochim. Acta* **55**, 193–209 (1991).
- Dunn, T. & Sen, C. Mineral/matrix partition coefficients for orthopyroxene, plagioclase, and olivine in basaltic to andesitic systems: a combined analytical and experimental study. *Geochim. Cosmochim. Acta* **58**, 717–733 (1994).
- Green, T. H. & Pearson, N. J. An experimental study of Nb and Ta partitioning between Ti-rich minerals and silicate liquids at high pressure and temperature. *Geochim. Cosmochim. Acta* **51**, 55–62 (1987).
- Jenner, G. A. *et al.* Determination of partition coefficients for trace elements in high-pressure-temperature experimental run products by laser ablation microprobe - inductively coupled mass spectrometry (LAM-ICP-MS). *Geochim. Cosmochim. Acta* **57**, 5099–5103 (1993).
- Klemme, S., Blundy, J. D. & Wood, B. J. Experimental constraints on major and trace element partitioning during partial melting of eclogite. *Geochim. Cosmochim. Acta* (in the press).
- Jacob, D. E. & Foley, S. F. Evidence for Archean oceanic crust with low high field strength element signature from diamondiferous eclogite xenoliths. *Lithos* **48**, 317–336 (1999).
- Rudnick, R. L., Barth, M., Horn, I. & McDonough, W. F. Rutile-bearing refractory eclogites: Missing link between continents and depleted mantle. *Science* **287**, 278–281 (2000).
- Stolz, A. J., Jochum, K. P., Spettel, B. & Hofmann, A. W. Fluid- and melt-related enrichment in the subarc mantle: evidence from Nb/Ta variations in island arc basalts. *Geology* **24**, 587–590 (1996).
- Panjasawatwong, Y., Danyushevsky, L. V., Crawford, A. J. & Harris, K. L. An experimental study of the effects of melt composition on plagioclase-melt equilibria at 5 and 10 kbar: implications for the origin of magmatic high-An plagioclase. *Contrib. Mineral. Petrol.* **118**, 420–432 (1995).
- Ionov, D. A. & Hofmann, A. W. Nb-Ta-rich mantle amphiboles and micas: implications for subduction-related metasomatic trace element fractionations. *Earth Planet. Sci. Lett.* **131**, 341–356 (1995).
- Scambelluri, M. *et al.* Incompatible element-rich fluids released by antigorite breakdown in deeply subducted mantle. *Earth Planet. Sci. Lett.* **192**, 457–470 (2001).
- Rollinson, H. R. Petrology and geochemistry of metamorphosed komatiites and basalts from the Sula Mountains greenstone belt Sierra Leone. *Contrib. Mineral. Petrol.* **134**, 86–101 (1999).

Acknowledgements

Discussions with R. Overt, M.G. Barth and D.E. Jacob helped to shape the findings reported here.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.F. (e-mail: sfoley@uni-greifswald.de).

Paired gill slits in a fossil with a calcite skeleton

Patricio Dominguez^{*†}, Antone G. Jacobson[‡] & Richard P. S. Jefferies^{*}

^{*} Department of Palaeontology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

[‡] Patterson Labs, MCDB, The University of Texas, Austin, Texas 78712, USA

[†] Departamento de Paleontología, Universidad Complutense de Madrid, Ciudad Universitaria, 28040 Madrid, Spain

The chordates, hemichordates (such as acorn worms) and echinoderms (such as starfish) comprise the group Deuterostomia, well established as monophyletic^{1,2}. Among extant deuterostomes, a skeleton in which each plate has the crystallographic structure of a single crystal of calcite is characteristic of echinoderms and is always associated with radial symmetry and never with gill slits. Among fossils, however, such a skeleton sometimes occurs without radial symmetry. This is true of *Jaekelocarpus oklahomensis*, from the Upper Carboniferous of Oklahoma, USA, which, being externally almost bilaterally symmetrical, is traditionally placed in the group Mitrata (Ordovician to Carboniferous periods, 530–280 million years ago), by contrast with the bizarrely asymmetrical Cornuta (Cambrian to Ordovician periods, 540 to 440 million years ago). Using computer X-ray microtomography, we describe the anatomy of *Jaekelocarpus* in greater detail than formerly possible, reveal evidence of paired gill slits internally and interpret its functional anatomy. On this basis we suggest its phylogenetic position within the deuterostomes.

Three widely divergent views of the fossils known as mitrates and cornutes are now current—one chordate interpretation^{3–8}, and a first^{9–13} and second^{14,15} echinoderm interpretation. The present paper redescribes the mitrate *J. oklahomensis* from the Upper Carboniferous Gene Autry Shale Formation near Ardmore, Oklahoma, USA, which was originally described under the second echinoderm interpretation¹⁵. We redescribe it under the chordate interpretation, initially as a convention so that useful words such as 'ventral' or 'posterior' shall be unambiguous.

Jaekelocarpus, like all other mitrates and also like an extant

tunicate tadpole, consists of a head and a tail. We reconstruct its external appearance in Figs 1 and 2.

The tail can be only tentatively reconstructed in *Jaekelocarpus* but, as in other mitrates, is made up of fore-, mid- and hind-tail. It probably served to pull the animal rearwards in life^{6,7,13} by the use of big muscles located in the fore tail.

The head was dorsally flat and ventrally convex and its skeleton was constructed of ten calcite plates only. These are given letters intended to imply homology with particular plates in other mitrates and cornutes⁴ (Fig. 1). The ventral skeleton would have been rigid and consists of plate θ on the left, ϵ on the right and π anteriorly. The externally visible dorsal skeleton consists of plate h at posterior right, i at posterior left, d at median anterior and c (oral spine) articulated to the anterior end of d . Also in the dorsal skeleton but inside the animal are plate e in contact with the right margin of d , plate k/l in contact with the left margin of d and plate μ joined to the ventral ends of e and k/l . Plates e , k/l and μ , together with the externally visible d , form a mouth frame not previously described. The mouth was located in a gap between d above and π below. There are two large antero-dorsal gaps in the skeleton, not previously observed. The left antero-dorsal gap was bordered by ventral plate θ laterally and by k/l and d medially while the right antero-dorsal gap was bordered by ventral plate ϵ laterally and plates d and e medially. These gaps are certainly not due to post-mortem damage because the plates in contact with them are round-edged rather than fractured.

Two complicated structures, here named the left and right slit-bearing complexes, are internal extensions of the ventral plates θ and ϵ respectively and are rigidly attached postero-dorsally, by a short isthmus of calcite, to the anterior surface of a low vertical wall on the internal surface of those plates. They are built of delicate imperforate stereom about 20 μ m or less in thickness. They have, on right and left, a dorsal, almost horizontal imperforate lamina (the 'roof' of the complex) which passes laterally into a ventro-medially sloping lamina (the 'side wall' of the complex) penetrated by at least three slits. These are arranged in a dorso-ventral series (L1, L2 and L3 on the left and R1, R2 and R3 on the right), are horizontally elongate with their dorsal edges almost parallel to the ventral edges and are each connected to a short calcite tube which extends medianwards approximately perpendicularly to the slit and with the same cross-section as the slit. The connection, at the slit, of each

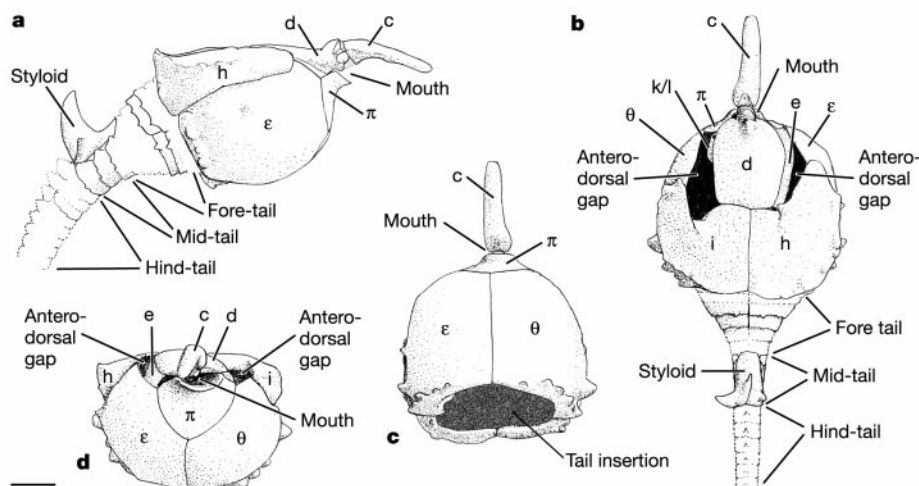


Figure 1 Reconstruction of *Jaekelocarpus oklahomensis* Kolata, Frest & Mapes 1991 (ref. 15) (Gene Autry Shale Formation, Upper Carboniferous, near Ardmore, Oklahoma, USA). The head is reconstructed by X-ray microtomography as shown in Fig. 2, and the tail is reconstructed on the basis of several plates observed in the paratypes. No attempt has been made to reconstruct the soft parts around the mouth nor in the antero-dorsal gaps,

but muscles which probably raised and lowered the oral spine c are indicated. For meaning of plate notation, see text. Scale bar, 1 mm. **a**, Right lateral aspect, tail curved downwards as was probably common in life. **b**, Dorsal aspect, tail straight. **c**, Ventral aspect of head. **d**, Anterior aspect of head.

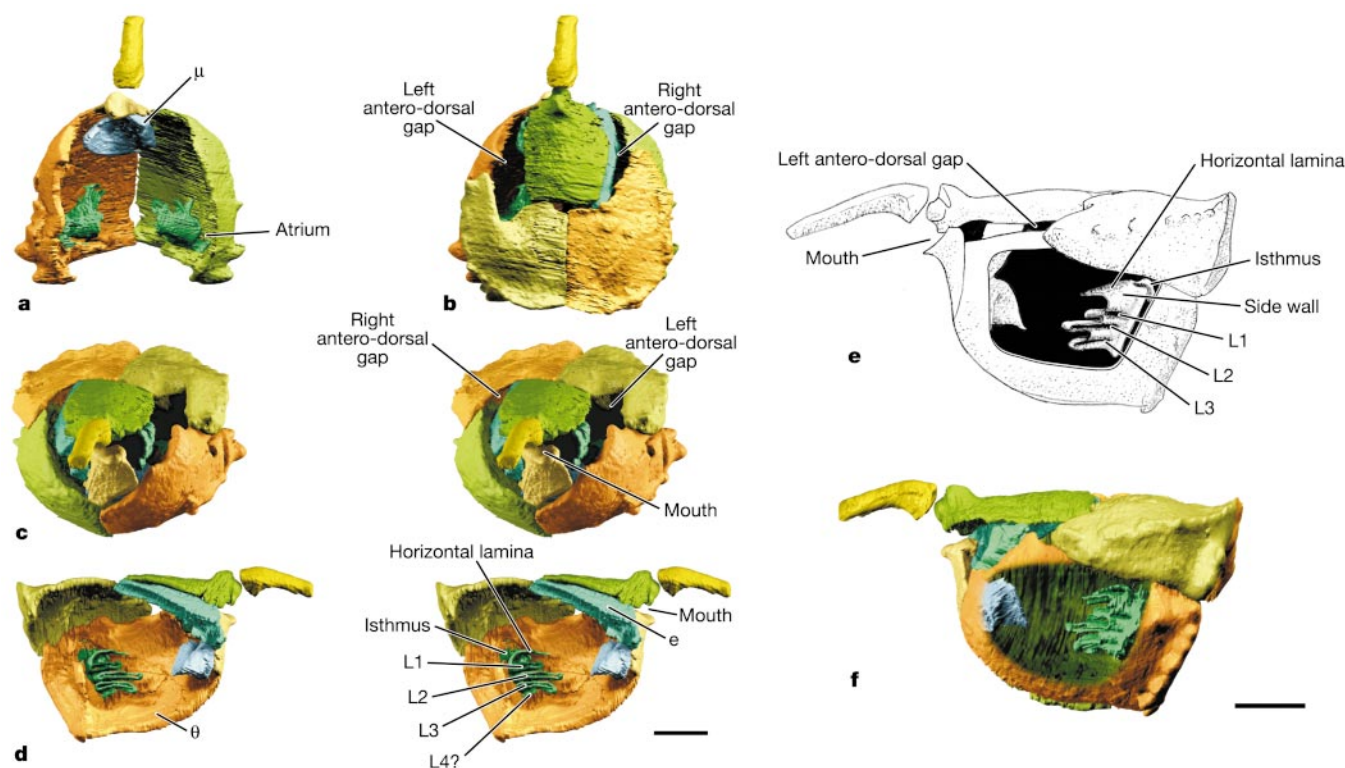


Figure 2 Microtomographic reconstructions of *J. oklahomensis* Burke Museum, University of Washington, UWBM74305. Plates shown in life position. For plate notation see Fig. 1 and text. The applied colouring of a plate distinguishes it from its neighbours, except for the blue-green colour of both slit-bearing complexes, which are parts of plates ϵ and θ . No attempt has been made to correct damage to the antero-dorsal edge of plate θ or the left and right edges of π . Plates μ , e and k/l are ventral plates of the mouth frame (previously unreported). Panels **c**, **d** are stereo pairs. **a**, Dorsal aspect of ventral skeleton showing internal features including dorsal surfaces of the paired slit-bearing complexes

(blue-green). **b**, Dorsal aspect of head. **c**, Antero-dorsal aspect to emphasise the size of the antero-dorsal gaps; π (pale yellow) is incomplete at right and left and ϵ (orange) is incomplete antero-dorsally. The blue plates belong to the internal mouth frame. **d**, The left slit-bearing complex (blue-green), attached to the rest of plate θ (orange), exposed in median aspect by removal of plates h and ϵ from the right side of the skeleton; the blue plates belong to the internal mouth frame. **e**, Explanatory drawing of **f**. **f**, Left slit-bearing complex, in lateral aspect, shown in relation to the rest of the skeleton by rendering the rest of plate θ transparent. Scale bars, 1 mm. (See also Supplementary Information.)

tube to the side wall of the slit-bearing complex is not abrupt but is everywhere rounded. In addition to the three pairs of slits already mentioned, there are, on left and right, possible indications of a fourth slit (L4, R4) ventral to the others. There is a clear space lateral to each slit-bearing complex, between it and the internal surface of the rest of plate ϵ on the right or θ on the left. Because of the rigid isthmus connecting the slit-bearing complex with the rest of the plate that bears it, this space is geometrically required by the skeletal anatomy, not reconstructional. Figure 3 shows the left and right complexes in their post-mortem positions.

The elongate shape of the slits and tubes and their remarkably constant, very low height measured in tens of micrometres (70 to 100 μm ; mean 75 μm) which is irrespective of the varying length (range 510–820 μm), strongly suggests that they were lined in life with ciliated epithelia functioning to pump water through the tubes (Table 1). A comparison can be drawn with the pumping ciliated interfilamentary slits in the gills of the bivalve *Mytilus edulis*, the width (height) of which is remarkably constant at about 40 μm (occupied by two opposed sets of 15 μm long cilia together with a

10 μm gap between the two sets), while the length varies but is some two orders of magnitude greater. The width of the slits is known to be physiologically critical in *Mytilus*, because flow through a slit decreases if the gap, normally 10 μm wide, between the two sets of cilia is diminished¹⁶. On the reasonable assumption that the calcitic slits of *Jaekelocarpus* were lined with epithelia 20 μm thick, say, the dimensional resemblance to the slits of *Mytilus* is remarkable. The main difference is that the slits of *Mytilus* are longer and more numerous, but this is presumably in part because *Mytilus* is much bigger than *Jaekelocarpus*.

The gill slits (stigmata) of ascidian tunicates are similarly ciliated¹⁷. A provisional statistical study of the size and shape of the stigmata of Australian ascidians based on the scaled, camera-lucida drawings of Kott^{18–20} shows that these stigmata range in width (height) from 9 to 60 μm with a mean of about 23 μm (Fig. 4). At least up to a length of about 1 mm, ascidian stigmata have a weak tendency to increase in width as the length increases and the slits of *Jaekelocarpus* are longer than those of most ascidians. In consequence, assuming that the calcitic slits were lined with epithelia

Table 1 Dimensions of the slits in micrometres

Left	Slit length	Mean height*	Tube length	Right	Slit length	Mean height*	Tube length
L1	510	75	450	R1	640	75	520
L2	820	75	380	R2	>380†	75	390
L3	720	75	280	R3	760	75	190

A possible fourth slit has been ignored.
* The range in height for the slits is 70–100 μm .
† Incomplete slit.

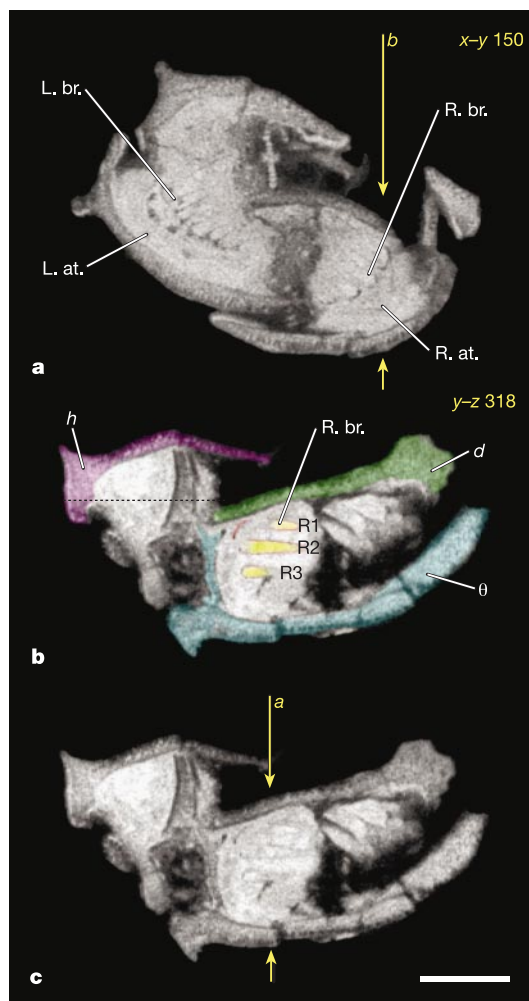


Figure 3 Microtomographic sections of *J. oklahomensis* Burke Museum, University of Washington, UWBM74305. **a**, Transverse section (original X-ray tomographic section no. 150, plane $x-y$) of left and right slit-bearing complexes in their post-mortem positions in the plates which bear them (θ to right, ϵ to left). The yellow arrows show the intersection of **b** and **c** with **a**. **b**, **c**, Secondary generated longitudinal section (no. 318, plane $y-z$); the yellow arrows in **c** show the intersection of **a** with **b** and **c**. **b**, Projection (yellow) of the median ends of the tubes of slits R1, R2 and R3 onto plane b of **a**, **c**. As **b**, but uncoloured, to show the objective evidence for the position of the branchial slits. Scale bar, 1 mm. L. at., left atrium; L. br., left branchial slit; R. at., right atrium; R. br., right branchial slit.

between 7.5 and 20 μm thick, they would fall naturally along the regression line calculated for the ascidians (Fig. 5).

As already mentioned, each slit has a short calcitic tube projecting inwards from it. Water probably flowed outwards, first through the tube and then through the slit. This is suggested by comparison with

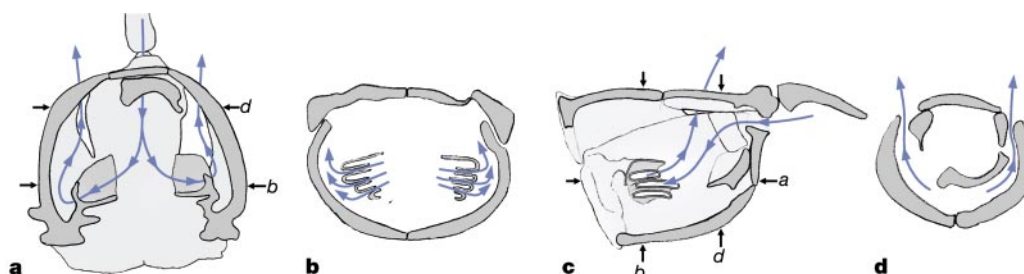


Figure 4 Suggested flow of water through the head of *Jaekelocarpus*. **a**, Ventral aspect. **b**, Transverse section through the slits. **c**, Sagittal section and left half of head. **d**, Transverse section through oral frame. Water (arrows show flow) enters through the

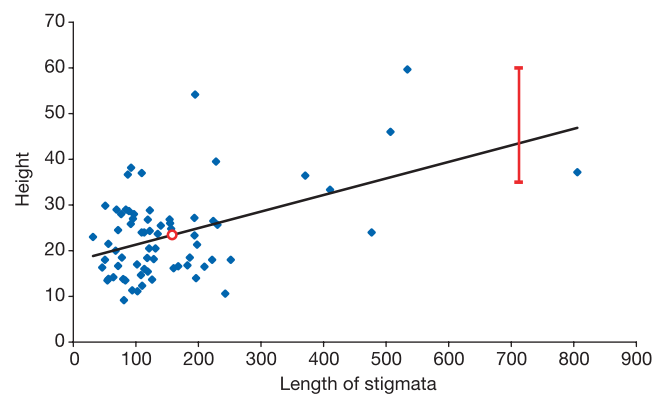


Figure 5 Length versus height (in μm), regression line (black) and mean (red circle) of ascidian stigmata (blue diamonds) in relation to the slits of *Jaekelocarpus* (vertical red line), assuming that the calcitic slits of the latter (mean height 75 μm) were lined with an epithelium between 7.5 and 20 μm thick, giving an estimated height between 35 and 60 μm . The data points (blue diamonds) for extant ascidian stigmata were taken from the scales drawings of Kott¹⁸⁻²⁰. Each point represents the mean value for a species (69 species; $n = 421$ stigmata).

the gill slits of cornutes (always on the left side of the head only) which are presumably broadly homologous with the left gill slits of tunicates and which often have a plausible outlet-valve structure^{3,21,22}.

After flowing outwards through a slit (Fig. 4) water probably entered an unpaired chamber (atrium), or a pair of chambers (atria), which would have filled the space visible in transverse sections lateral to the slit-bearing complexes but internal to the rest of plates ϵ and θ . This chamber or chambers probably debouched through the large antero-dorsal gaps left and right of plate d , which would otherwise have no obvious function.

The slits of *Jaekelocarpus* have been suggested to be respiratory¹⁵, and perhaps gill slits²¹. These suggestions were made, however, on the basis of the broken material studied and drawn by Kolata *et al.*¹⁵. X-ray microtomography strengthens the comparison with gill slits by revealing the anterior boundaries of the slits for the first time and yielding dimensions consistent with the presence of a ciliary pump. This functional analysis suggests two general conclusions:

(1) The slits in the slit-bearing complexes of *Jaekelocarpus* were probably ciliated water-pumping gill slits, passing water outwards. Water passing out from the slits would enter the spaces (atria) lateral to the slits and, probably, would escape from the animal through the antero-dorsal gaps described here for the first time. Among extant chordates, such ciliated slits occur in acraniates (for example, amphioxus = *Branchiostoma*) and tunicates.

(2) *Jaekelocarpus* was probably a tunicate because it shows several potential synapomorphies with extant tunicates. These are that the probably ciliated gill slits were antero-posteriorly elongate like the

mouth, passes through the slits at right and left (see **a**, **b**, **c**) into the left and right atria (see **b**, **c**) and leaves the head through the antero-dorsal gaps (see **a**, **c**, **d**).

stigmata of ascidian tunicates²³, arranged in paired dorso-ventral columns like the stigmata of ascidian and doliolid tunicates²³, while the presumed atria extended far forwards right and left of the stigmata and therefore right and left of the pharynx, as in ascidian tunicates²³. Furthermore, the atria opened antero-dorsally as in ascidian tunicates²³, and these anterior openings were paired as in post-larval ascidian tunicates²⁴.

As a tunicate, *Jaekelocarpus* probably belonged to the stem group of the Tunicata because it retained the primitive features of a calcite skeleton and a downward-flexing tail which, as parsimony suggests, would not have existed in the latest common ancestor of extant tunicates. If a tunicate, *Jaekelocarpus* would necessarily belong to the crown group of the chordates, as the chordate interpretation of mitrates and cornutes has long proposed for mitrates in general^{22,25}. □

Methods

The technique we used is high-resolution X-ray computed tomography (μ CT). It is non-destructive and permits the study of non-transparent objects such as fossils, producing images that correspond to serial sections with resolution as good as 8 μ m. The sections were used to create three-dimensional computer models of the objects by reconstructing the surfaces that connected corresponding outlines on adjacent sections^{26,27}.

The specimen of *Jaekelocarpus oklahomensis* chosen for complete scanning (Burke Museum, University of Washington, UWBM74305) was selected after preliminary tests on a total of three specimens. The scanning was carried out, under the direction of T. Rowe, at the University of Texas High-Resolution X-Ray CT Facility, Department of Geological Sciences, USA. The parameters of the scan were: 150 kV; 0.053 mA; slice thickness 0.016 mm; interslice spacing 0.016 mm; diameter of field of view 4.53 mm. 276 sections were saved as 16-bit TIFF files. The matrix data size was 512 $\times$ 512 $\times$ 376 voxels and the voxel size was 8.84 $\times$ 8.84 $\times$ 16 μ m (see Supplementary Information).

The individual plates were recognized, and distinguished from each other as separate entities (segmented), using Mimics 6.3 software (Materialise N. V.). In Mimics, the data can be explored in three views: the original images in x - y planes and resliced images in x - z or y - z planes. Because of lateral changes in the X-ray attenuation values within individual skeletal plates, it was necessary to use local thresholds.

The CTM software module of Mimics interpolates the slice data to generate three-dimensional STL files, one for each plate or anatomical region. Such STL files were exported to Rhinoceros 1 (Robert McNeel) as polygon meshes, where models of the plates were moved in virtual space and so placed in the original life positions relative to each other. To prevent artefacts after these manipulations, the size parameters of the plates (length and volume) were checked in Communicator 1.6 (Materialise N. V.) by means of which we obtained secondary serial sections of the restored specimen.

Received 26 July 2001; accepted 12 March 2002; doi:10.1038/nature00805.

- Peterson, K. J. & Eernisse, D. J. Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA. *Evol. Dev.* **3**, 170–205 (2001).
- Cameron, C. B., Garey, J. R. & Swalla, B. J. Evolution of the chordate body plan: new insights from phylogenetic analyses of deuterostome phyla. *Proc. Natl Acad. Sci. USA* **97**, 4469–4474 (2000).
- Jefferies, R. P. S. Some fossil chordates with echinoderm affinities. *Symp. Zool. Soc. Lond.* **20**, 163–208 (1967).
- Jefferies, R. P. S. In *Major Events in Vertebrate Evolution* Ch. 4 (ed. Ahlberg, P. E.) 20–66 (Systematics Association and Taylor & Francis, London, 2000).
- Jefferies, R. P. S. & Jacobson, A. G. An episode in the ancestry of the vertebrates: from mitrate to crown-group craniate. *Integr. Biol.* **1**, 115–132 (1998).
- Sutcliffe, O. E., Südkamp, W. H. & Jefferies, R. P. S. Ichnological evidence on the behaviour of mitrates: two trails associated with the Devonian mitrate. *Rhenocystis. Lethaia* **33**, 1–12 (2000).
- Jefferies, R. P. S. Ichnological evidence on the behaviour of mitrates: comments on the reply. *Lethaia* **34**, 262 (2001).
- Cripps, A. P. A new stem-group chordate (Cornuta) from the Llandeilo of Czechoslovakia and the cornute-mitrate transition. *Zool. J. Linn. Soc.* **96**, 49–85 (1989).
- Ubahgs, G. In *Treatise on Invertebrate Paleontology Part 5. Echinodermata* (ed. Moore, R. C.) S297–S650 (Kansas Univ. Press and Geological Society of America, New York, 1968).
- Parsley, R. L. In *Echinoderm Phylogeny and Evolutionary Biology* (eds Paul, C. R. C. & Smith, A. B.) 347–361 (Oxford Scientific Publications/Liverpool Geological Society, Oxford/Liverpool, 1988).
- Ruta, M. Brief review of the stylophoran debate. *Evol. Dev.* **1**, 123–135 (1999).
- David, B., Lefebvre, B., Mooi, R. & Parsley, R. L. Are homalozoans echinoderms? An answer from the extralaxial-axial theory. *Paleobiology* **26**, 529–555 (2000).
- Lefebvre, B. & David, B. Ichnological evidence on the behaviour of mitrates: reply. *Lethaia* **34**, 260–261 (2001).
- Philip, G. M. Carpoids: echinoderms or chordates? *Biol. Rev.* **54**, 439–471 (1979).
- Kolata, D. R., Frest, T. J. & Mapes, R. H. The youngest carpod: occurrence, affinities and life mode of a Pennsylvanian (Morrowan) mitrate from Oklahoma. *J. Paleontol.* **65**, 844–855 (1991).
- Jørgensen, C. B. Water processing in ciliary feeders, with special reference to the bivalve filter pump. *Comp. Biochem. Physiol.* **94a**, 383–394 (1989).
- Martinucci, G. B., Burighe, P. & Dallai, R. In *Form and Function in Zoology* (eds Lanzavecchia, G. & Valvassori, R.) Vol. 5, 123–140 (Selected Symposia and Monographs, Unione Zoologica Italiana, Mucchi, Modena, 1991).
- Kott, P. The Australian Ascidiacea. Part 1. Phlebobranchia and Stolidobranchia. *Memoirs Queensland Mus.* **20**, 1–47 (1985).
- Kott, P. The Australian Ascidiacea. Part 2. Aplousobranchia. *Memoirs Queensland Mus.* **29**, 1–298 (1990).

- Kott, P. The Australian Ascidiacea. *Memoirs Queensland Mus.* **32**, 375–620 (1992).
- Jefferies, R. P. S. A defence of the calcichordates. *Lethaia* **30**, 1–10 (1997).
- Jefferies, R. P. S. *The Ancestry of the Vertebrates* (Cambridge Univ. Press and British Natural History Museum, London, 1986).
- Berrill, N. J. *The Tunicata with an Account of the British Species* (Ray Society, London, 1950).
- Huus, J. Genitalorgane und Ganglio-Genital-Strang bei *Corella parallelogramma* O. F. M. *Skrifter. Videnskabselskabet Christiania*. **19**, 1–50 (1924).
- Jefferies, R. P. S. In defence of the calcichordates. *Zool. J. Linn. Soc.* **73**, 351–396 (1981).
- Ketcham, R. A. & Carlson, W. D. Acquisition, optimization and interpretation of X-ray computed tomographic imagery: applications to the geosciences, three-dimensional reconstruction, modeling and visualization of geological materials. *Comp. Geosci.* **27**, 381–400 (2001).
- van Geet, Swennen, R. & Wevers, M. Quantitative analysis of reservoir rocks by microfocus X-ray computerized tomography. *Sedim. Geol.* **132**, 1–2 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>). The original TIFF files showing the serial X-ray sections of *Jaekelocarpus* are also stored in the Digital Morphology site of the University of Texas and can be accessed on <http://www.digimorph.org>.

Acknowledgements

We are grateful to E. Nesbitt and R. Eng for making the specimens of *Jaekelocarpus* available and to T. Rowe and his co-workers for their collaboration. P.D. was supported on a post-doctoral fellowship in London by the Spanish Government (Ministerio de Educación, Cultura y Deportes).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.P.S.J. (e-mail: r.jefferies@nhm.ac.uk).

Positive interactions among alpine plants increase with stress

Ragan M. Callaway*, R. W. Brooker†, Philippe Choler‡, Zaal Kikvidze§, Christopher J. Lortie||, Richard Michalet¶, Leonardo Paolini#, Francisco I. Pugnaire☆, Beth Newingham*, Erik T. Aschehoug*, Cristina Armas☆, David Kikodze§ & Bradley J. Cook*

* Division of Biological Sciences, University of Montana, Missoula, Montana 59812, USA

† Centre for Ecology and Hydrology, CEH Banchory Research Station, Hill of Brathens, Banchory, AB31 4BY, UK

‡ Station Alpine du Lautaret et Laboratoire de Biologie des Populations d'Altitude UMR CNRS-UJF 5553, University Joseph Fourier, Grenoble, BP 53, 38041 Grenoble cedex, France

§ Institute of Botany of the Georgian Academy of Sciences, Kojori Road 1, Tbilisi 380007, Georgia

|| Department of Botany, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada

¶ Ecosystèmes et Changements Environnementaux, University Joseph Fourier, Grenoble, BP 53, 38041 Grenoble cedex, France

Laboratorio de Investigaciones Ecológicas de las Yungas, UNT, CC34, Yerba Buena, Tucuman, Argentina

☆ Estación Experimental de Zonas Áridas, Consejo Superior de Investigaciones Científicas, General Segura 1, 04001 Almería, Spain

Plants can have positive effects on each other¹. For example, the accumulation of nutrients, provision of shade, amelioration of disturbance, or protection from herbivores by some species can enhance the performance of neighbouring species. Thus the notion that the distributions and abundances of plant species are independent of other species may be inadequate as a theoretical underpinning for understanding species coexistence and

** Present address: The Nature Conservancy, 201 Mission Street, San Francisco, California 94105, USA.

Table 1 Descriptions of experimental sites

Site	Latitude, longitude	Low-high elevations (m)	Low-high microtopography	Set-up-harvest dates	Low-high species number	Measurements
Central Brooks Range, Alaska	68.1, 211.0	800–1,400	Flat-weak concave	1998–99	5, 5	Leaf number & biomass
Abisko, Sweden	68.2, 18.5	580–1,000	Concave-concave	1998–99	3, 4	Leaf number & biomass
Kluane Range, Yukon	60.4, 221.9	900–1,750	Weak concave-Concave	1998–99	5, 5	Leaf number & biomass
Cairngorms, Scotland	57.1, 3.5	400–740	Concave-flat	2000	5, 5	Leaf number & biomass
Rocky Mountains, Banff	51.3, 244.0	1,400–2,300	Concave-convex	2000	4, 4	Leaf number & biomass
Rocky Mountains, Absaroka Mountains	45.1, 250.8	2,350–3,000	Flat-flat	1998–99	5, 5	Leaf number & biomass
French Alps	44.5, 6.4	2,100–2,900	Concave-convex	1997–98	10, 10	Leaf number & biomass
Central Caucasus, Kazbegi	42.5, 44.4	2,100–3,000	Flat-flat	1996–97	7, 5	Leaf number
Rocky Mountains, Colorado	40.2, 254.6	2,930–3,500	Flat-flat	1998–99	5, 5	Leaf number & biomass
Sierra Nevada, Spain	37.1, 3.4	2,400–3,100	Flat-weak concave	1998–99	4, 4	Leaf number & biomass
Central Andes, Tucuman	26.5, 294.9	2,000–3,600	Flat-flat	1997–98	5, 5	Leaf number

Species are presented in the Supplementary Information.

diversity². But there have been no large-scale experiments designed to examine the generality of positive interactions in plant communities and their importance relative to competition. Here we show that the biomass, growth and reproduction of alpine plant species are higher when other plants are nearby. In an experiment conducted in subalpine and alpine plant communities with 115 species in 11 different mountain ranges, we find that competition generally, but not exclusively, dominates interactions at lower elevations where conditions are less physically stressful. In contrast, at high elevations where abiotic stress is high the interactions among plants are predominantly positive. Furthermore, across all high and low sites positive interactions are more important at sites with low temperatures in the early summer, but competition prevails at warmer sites.

The performance of plants in communities is very different from

their performance as individuals^{3,4}, a discrepancy that is generally attributed to negative, competitive interactions^{5,6}. However, recent research has demonstrated that the overall effects of species on each other may vary as competition and facilitative mechanisms shift in relative importance^{1,2,7,8}. In an attempt to integrate biotic interactions and abiotic factors into a single conceptual model, ecologists have hypothesized that the relative importance of competition and facilitation may vary inversely along gradients of abiotic stress^{8,9}. The importance of facilitation as an organizing process in communities is predicted to increase along gradients of increasing abiotic stress and decreasing productivity, whereas the importance of competition is predicted to decrease. We tested the hypothesis that the importance of facilitation, relative to competition, increases in importance in response to increasing abiotic stress in a global experiment in subalpine and alpine plant communities.

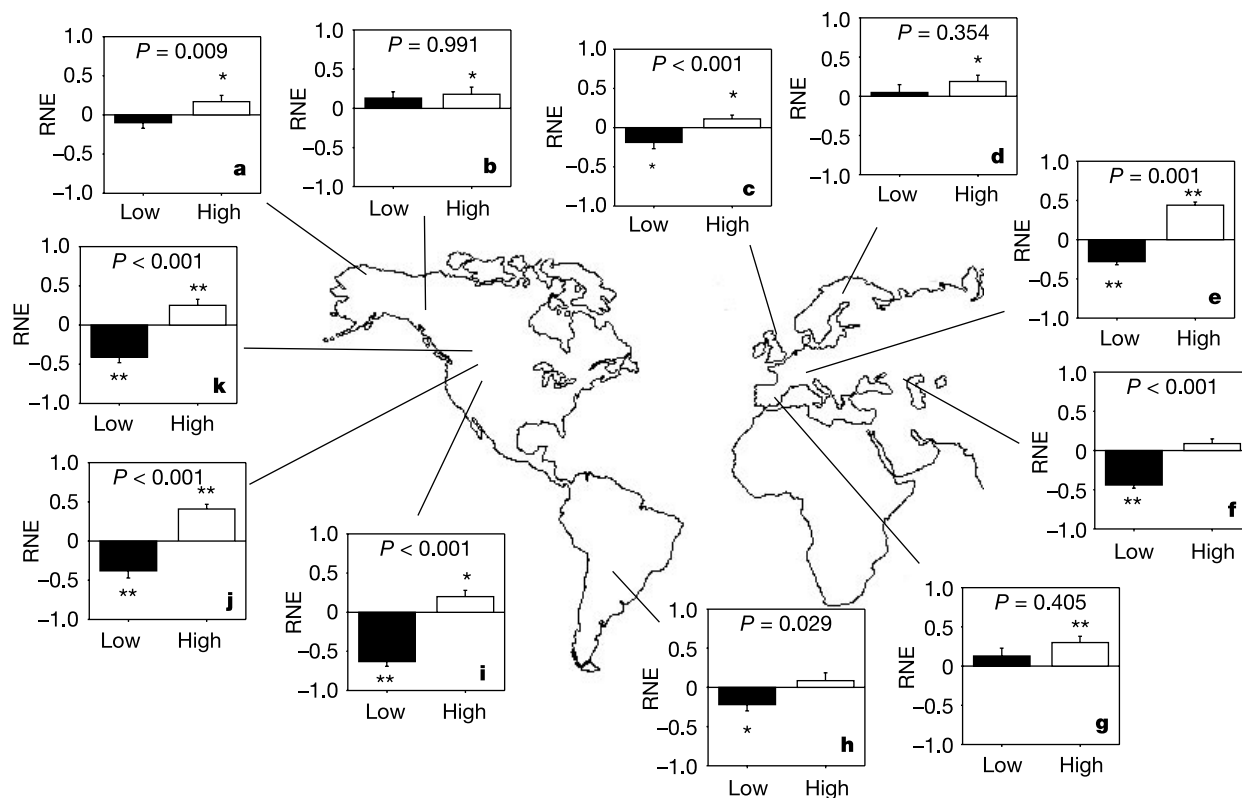


Figure 1 Relative neighbour effect (RNE) at the 11 experimental sites. Error bars represent one standard error, P values denote significance of differences between the two sites (ANOVA with site and species as main effects), and asterisks denote a site effect that was significantly different from zero ($P > 0.01$). RNE was calculated for biomass of target plants at all sites except for the Caucasus and the Andes sites, where RNE was calculated

using the difference in leaf number between control and removal target pairs. **a**, Brooks Range; **b**, Kluane Range; **c**, Cairngorms; **d**, Abisko; **e**, French Alps; **f**, Caucasus Mountains; **g**, Sierra Nevada; **h**, Andes; **i**, Rocky Mountains, Colorado; **j**, Absaroka Mountains, Montana; **k**, Rocky Mountains, Banff.

We compared the effects of removing neighbouring vegetation on 8–12 replicates of 115 different target plant species at low versus high elevations in 11 different mountain ranges around the world (Table 1). Over all locations combined, plant–plant interactions shifted from competition in relatively benign abiotic environments at low elevations to facilitation in more stressful environments at high elevations: Fig. 1; analysis of variance (ANOVA) for relative neighbour effect (RNE)¹⁰ for biomass (RNE_{biomass}), $F_{\text{global location}} = 1.22$, degrees of freedom, d.f. = 8, 98, $P = 0.293$; $F_{\text{site altitude}} = 51.06$, d.f. = 1, 98, $P < 0.001$; ANOVA for $RNE_{\text{leaf growth}}$, $F_{\text{global location}} = 1.34$, d.f. = 9, 93, $P = 0.262$; $F_{\text{site altitude}} = 59.97$, d.f. = 1, 93, $P < 0.001$. For the nine locations where plants were harvested, RNE_{biomass} was -0.22 ± 0.02 (1 s.e.) at the low sites and $+0.25 \pm 0.02$ at the high sites. For the nine locations where leaf growth rates were measured $RNE_{\text{leaf growth}}$ was 0.33 ± 0.02 (1 s.e.) at the low sites and $+0.16 \pm 0.02$ at the high sites. At eight of the 11 locations, removal of neighbours had effects on aboveground target plant biomass that were significantly different, and more facilitative, at high elevations than at low elevations, indicating that general interactions shifted from competition to facilitation with increasing elevation and abiotic stress. At the other three sites RNE was significantly greater than zero, indicating facilitation, but because RNE was also positive at the low sites there was no difference between sites. Two of these three sites were in highly stressful arctic environments (the Kluane Mountains in western Canada and Abisko, Sweden), and the third was in the Sierra Nevada of Spain where both high and low sites were unusually dry. No significant effect of competition on plant biomass (for all species combined) was found at any of the high-elevation experimental sites. For all species combined across all 11 locations, neighbours had weakly competitive effects on mortality of target plants at low elevations, but highly facilitative effects on mortality at high elevations (Fig. 2). Similarly, sexual reproduction was reduced by neighbours at low elevations and enhanced by neighbours at high elevations (Fig. 3). Considered together, the results of this short-term experiment demonstrate strong shifts from competitive processes in relatively benign environments to facilitation in more stressful environments as biological determinants of plant reproduction, community composition and community diversity.

We hypothesize that the shift from competition at low elevations to facilitation at high elevations is based on fundamental physiological limitations. As conceptualized by Grime¹¹ for shifts in competitive intensity on stress gradients, we believe that non-resource factors such as temperature, wind, and soil disturbances are less limiting to plant growth at low elevations, permitting plants to grow to the point where further growth or reproduction is limited

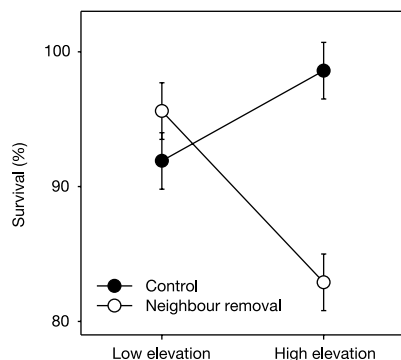


Figure 2 Proportion of surviving target species in controls and neighbour removal treatments at high and low elevation experimental sites for all 11 locations combined. Error bars show one standard error. (ANOVA, $F_{\text{treatment}} = 8.47$, d.f. = 1, 206, $P = 0.004$; $F_{\text{site}} = 2.10$, d.f. = 1, 206, $P = 0.149$; $F_{\text{treatment} \times \text{site}} = 22.13$, d.f. = 1, 206, $P < 0.001$.)

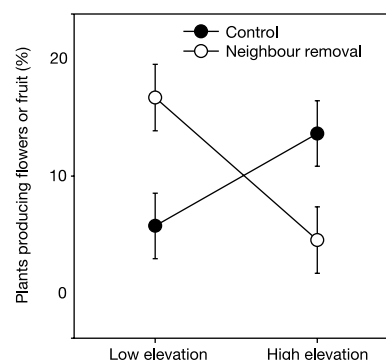


Figure 3 Proportion of flowering or fruiting target species in controls and neighbour removal treatments at high and low elevation experimental sites for all 11 locations combined. Error bars show one standard error. (ANOVA, $F_{\text{treatment}} = 0.62$, d.f. = 1, 180, $P = 0.432$; $F_{\text{site}} = 0.14$, d.f. = 1, 180, $P = 0.709$; $F_{\text{treatment} \times \text{site}} = 16.17$, d.f. = 1, 180, $P < 0.001$.) The degrees of freedom are lower than for survival because reproduction was not recorded at all sites.

by resources. At high elevations, temperature, wind scouring or soil instability may limit plant growth more than resource availability. Amelioration of these severe stresses by neighbours may favour growth more than competition for resources with the same neighbours impairs growth.

Experimental studies of facilitation and competition seldom provide unbiased, neutral estimates of the relative occurrence or importance of these interactions in communities because focal species are chosen on the basis of their spatial association with other species. We made no effort to choose species that showed particular spatial relationships with any other species or that occupied any particular position on elevational or topographic gradients. Our selection of a large proportion of the species occurring in these alpine communities provides more support for generality of the importance of facilitation than studies that have focused on few species at one or two locations. Despite the problem of scaling our experiments to the longer time frame of alpine plant community development, our demonstration of common, strong facilitation supports the general neutral model constructed by Dodds¹² for community organization, and the empirical model developed by Miller to incorporate both direct and indirect effects among plants⁴. They found that positive interactions among species were as likely as negative ones in communities as long as relatively large numbers of species and connections were considered.

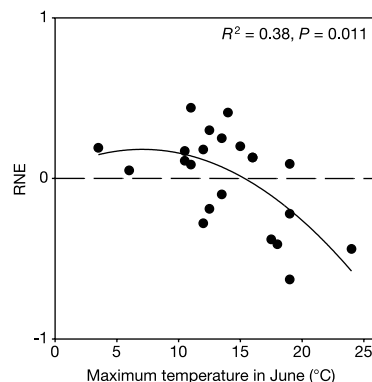


Figure 4 The relationship between relative neighbour effect (RNE) and the estimated maximum temperature in early summer (June in the Northern Hemisphere and December in the Southern Hemisphere) at each of the 22 experimental sites. Points above zero on the RNE axis indicate facilitation, whereas points below zero indicate competition.

Although the global pattern of our experiment provides compelling evidence for generality in the shift from competition to facilitation with increasing abiotic stress, the overall effect of neighbours on target species varied substantially among the geographical locations where we conducted experiments. Regression analyses of different combinations of temperature and precipitation variables with RNE values yielded only one significant relationship, a negative correlation between RNE and maximum June (December in the Andes) temperatures estimated for each of the 22 sites (Fig. 4). This correlation has important implications for predicting the response of alpine plant communities to climate change. Increased temperatures may alter the current balance of facilitation and competition in alpine plant communities and drive more rapid changes in composition and diversity than predicted by physiology-based models. Although we can only speculate about the relationship we found between the strength and direction of plant interactions and climate, understanding the effects of community-scale processes is crucial for predicting the responses of natural systems to global climate change. Studies of the effects of global warming on communities generally consider only the responses of individual plants or changes in community composition to simple environmental manipulations, and not experiments on the interactions between individuals^{13–15}.

As indicated by many other smaller-scale experiments, our large-scale experiment suggests that facilitation has been underestimated as a determinant factor in the organization and diversity of plant communities. This underestimation has consequences for fundamental community theory. The 'individualistic' paradigm of plant community organization is a fundamental tenet in ecology that emphasizes "the fluctuating and fortuitous immigration of plants and an equally fluctuating and variable environment"¹⁶. With the exception of competitive interactions, the individualistic theory asserts that the distributions and abundances of plant species are independent of other species^{16–18}. Individualistic theory has ramifications beyond academic boundaries. For example, the view that plant species are fully individualistic and 'interchangeable' in communities has been used to advocate active management towards "shaping and synthesizing new ecosystems, even in the 'natural' environment"¹⁹. The rationale for individualistic communities is based largely on correlative gradient analyses, in which species are almost always distributed independently of one another along environmental gradients^{20–22}. In other words, the fact that the distributions of species rarely overlap completely in nature has been interpreted as a lack of interdependence among species. In our experiment, plants demonstrated species-specific responses to the removal of the surrounding community, but the design of our experiment did not examine species-specific effects of neighbours. Therefore, our results cannot address interdependence as a species-specific phenomenon²³. However, our experimental results support a growing body of experimental evidence for frequent, strong positive interactions and interdependence within plant communities^{1,2,10–12}, and provide broad-based support for a predominant role of facilitation in plant communities in physically harsh environments. □

Methods

We chose two experimental sites at each of 11 different locations around the world (Table 1, Fig. 1). At each location one site was placed in subalpine herbaceous vegetation and the other was placed from 300 m (Cairngorms) to 1,200 m (Andes) higher in alpine vegetation. This allowed us to explore interaction strengths and directions along one of the complex abiotic gradients, elevation, thought to be the most important for alpine communities²⁴. We adopted Grime's perspective that productivity and biomass are correlated with stress⁵. The gradient from subalpine meadow to alpine tundra is correlated with large decreases in productivity and biomass because of lower temperatures and shorter growing seasons^{24,25}, and at all 11 locations in our study vegetation cover, biomass, and height were noticeably lower at the high sites than at the low sites.

Interactions among plants were assessed by the removal of all neighbouring species within 10 cm of a target individual, and comparison of the performance of target plants with neighbours removed to that of control target plants around which neighbours were

left intact. At each site 3–10 target species were chosen. We chose target individuals that were small relative to nearby conspecifics, and for which relatively distinct individuals or ramets could be found. Some of the target species were clonal, and to reduce the effect of clonality individuals within 10 cm of a conspecific were not used as targets. An analysis of non-clonal species at the experimental sites in the Alps yielded virtually the same results as the analysis for all species combined. Over all 11 of the locations a total of three species were not included in the experiments because discrete individuals were too difficult to find. For each species 8–12 pairs were chosen and one of each was randomly selected for neighbour removal.

We established removal experiments at the beginning of growing seasons from 1996 through 1999. Plants were harvested at the end of the following (second) growing season with the exception of the Cairngorms and the Banff locations where the experiment lasted only one growing season. At the time we established the experiments we counted the number of leaves for each target and control individual. At the end of the experimental periods we re-counted the number of leaves, counted flowers and fruits, recorded survival, harvested all aboveground parts of targets and controls, and measured their mass after oven drying for 3 days at 70 °C. Only the change in leaf number is presented for the Caucasus and Andes locations because of problems encountered collecting biomass; however, the results for change in leaf number were very similar to those for biomass at all locations. All statistical analyses were carried out using JMPin 4.0.2 software²⁶. We used modification of the relative competitive intensity (RCI) index where $RCI = (X_t - X_c)/X_t$. X is an estimation of plant performance in the presence (c) or in the absence (t) of neighbours. Because RCI is not symmetrical around zero we used a modified version of RCI, the "relative neighbour effect (RNE)"¹⁰. $RNE = (X_t - X_c)/x$ where x is the highest value of (X_t ; X_c). RNE ranges from -1 to +1 with negative values indicating facilitation and positive values competition. We present our results in the reverse, with positive values indicating facilitation and negative values competition for intuitive interpretation (that is, a positive bar equals a positive effect). In our experiments, $RNE_{biomass}$ was calculated as: (biomass of treated - biomass of control)/biomass of treated. For 20 of the replicates (about 2% of the total replicates) we used the biomass of the control (rather than the treatment) as the denominator in order to keep all RNE values between 1 and -1. RNE_{leaf} was calculated as: (final leaf number/initial leaf number of treated - final leaf number/initial leaf number of control)/(final leaf number/initial leaf number of treated). In 16 cases we used the final leaf number/initial leaf number of control as the denominator in order to keep all RNE values between 1 and -1. Differences between treated and control performance were computed for each pair of targets at a site. When any plant in a pair died the complete pair was excluded from the calculations. This approach provided a conservative estimate of plant-plant interactions, standardized across species, but retained the magnitude of variation within a species.

Using $RNE_{biomass}$ and RNE_{leaf} as dependent variables, we performed multiple ANOVAs with site altitude (low versus high) and geographical location, using the mean for all individuals within species to avoid pseudoreplication. Data were normally distributed. Survival of targets and occurrence of reproductive parts (flowers or fruits) at time of harvest were analysed by a log-linear analysis (JMPin 4.0.2).

Methodology for temperature analysis is presented in the Supplementary Information.

Received 20 February; accepted 25 March 2002; doi:10.1038/nature00812.

- Callaway, R. M. Positive interactions among plants. *Bot. Rev.* **61**, 306–349 (1995).
- Callaway, R. M. Positive interactions in plant communities and the individualistic-continuum concept. *Oecologia* **112**, 143–149 (1997).
- Berlow, E. L. Strong effects of weak interactions in ecological communities. *Nature* **398**, 330–334 (1999).
- Miller, T. E. Direct and indirect species interactions in an early old-field plant community. *Am. Nat.* **143**, 1007–1025 (1994).
- Grime, J. P. Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *Am. Nat.* **111**, 1169–1194 (1977).
- Connell, J. H. On the prevalence and relative importance of interspecific competition: evidence from field experiments. *Am. Nat.* **122**, 661–696 (1983).
- Pugnaire, F. I. & Luque, M. T. Changes in plant interactions along a gradient of environmental stress. *Oikos* **93**, 42–49 (2000).
- Brooker, R. W. & Callaghan, T. V. The balance between positive and negative plant interactions and its relationship to environmental gradients: a model. *Oikos* **81**, 196–207 (1998).
- Bertness, M. D. & Callaway, R. M. Positive interactions in communities. *Trends Ecol. Evol.* **9**, 191–193 (1995).
- Markham, J. H. & Chanway, C. P. Measuring plant neighbor effects. *Funct. Ecol.* **10**, 548–549 (1996).
- Grime, J. P. *Plant Strategies and Vegetation Processes* (Wiley, Chichester, 1979).
- Dodds, W. K. Interspecific interactions: constructing a general neutral model for interaction type. *Oikos* **78**, 377–383 (1997).
- Woodward, F. I., Smith, T. M. & Emanuel, W. R. A global primary productivity and phytogeography model. *Glob. Biogeochem. Cycles* **9**, 471–490 (1995).
- Shaver, G. R. & Jonasson, S. Response of arctic ecosystems to climate change: result of long-term field experiments in Sweden and Alaska. *Polar Res.* **18**, 245–256 (1999).
- Prentice, I. C. *et al.* A global biome model based on plant physiology and dominance, soil properties, and climate. *J. Biogeogr.* **19**, 117–134 (1992).
- Gleason, H. A. The individualist concept of the plant association. *Bull. Torrey Bot. Club* **53**, 7–27 (1926).
- Whittaker, R. H. *Communities and Ecosystems* (Macmillan, New York, 1975).
- Begon, M., Harper, J. L. & Townsend, C. R. *Ecology* 2nd edn 626–628 (Blackwell, London, 1990).
- Johnson, H. B. & Mayeux, H. S. Viewpoint: a view on species additions and deletions and the balance of nature. *J. Range Manag.* **45**, 322–333 (1992).
- Curtis, J. T. *The Vegetation of Wisconsin* (Univ. Wisconsin Press, Madison, 1959).
- Whittaker, R. H. A consideration of climax theory: the climax as population and pattern. *Ecol. Monogr.* **23**, 41–78 (1953).
- Austin, M. P. Continuum concept, ordination methods, and niche theory. *Annu. Rev. Ecol. Syst.* **16**, 39–61 (1985).

23. Callaway, R. M. Are positive interactions species-specific? *Oikos* **82**, 202–209 (1998).
24. Choler, P., Michalet, R. & Callaway, R. M. Facilitation and competition on gradients in alpine plant communities. *Ecology* **82**, 3295–3308 (2001).
25. Archibald, O. W. *Ecology of World Vegetation* 280–318 (Chapman and Hall, London, 1995).
26. JMPin 4.0.2 (SAS Institute Inc., Duxbury Press, Cary, North Carolina, 2000).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements

We thank the National Center for Ecological Synthesis and Analysis, The National Geographic Society, the Civilian Research and Development Foundation, and the Andrew W. Mellon Foundation for financial support.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.M.C. (e-mail: callaway@selway.umd.edu).

Consumer versus resource control of species diversity and ecosystem functioning

Boris Worm^{*†}, Heike K. Lotze^{*†}, Helmut Hillebrand[‡] & Ulrich Sommer[‡]

^{*} Biology Department, Dalhousie University, Halifax, Nova Scotia, B3H 4J1, Canada

[†] Section of Marine Ecology, Institute for Marine Science, Kiel University, Diesternbrooker Weg 20, 24105 Kiel, Germany

[‡] Erken Laboratory, Department of Limnology, Uppsala University, Norr Malma 4200, 76173 Norrtälje, Sweden

A key question in ecology is which factors control species diversity in a community^{1–3}. Two largely separate groups of ecologists have emphasized the importance of productivity or resource supply, and consumers or physical disturbance, respectively. These variables show unimodal relationships with diversity when manipulated in isolation^{4–8}. Recent multivariate models^{9–10}, however, predict that these factors interact, such that the disturbance–diversity relationship depends on productivity, and vice versa. We tested these models in marine food webs, using field manipulations of nutrient resources and consumer pressure on rocky shores of contrasting productivity. Here we show that the effects of consumers and nutrients on diversity consistently depend on each other, and that the direction of their effects and peak diversity shift between sites of low and high productivity. Factorial meta-analysis of published experiments confirms these results across widely varying aquatic communities. Furthermore, our experiments demonstrate that these patterns extend to important ecosystem functions such as carbon storage and nitrogen retention. This suggests that human impacts on nutrient supply¹¹ and food-web structure^{12,13} have strong and interdependent effects on species diversity and ecosystem functioning, and must therefore be managed together.

The most striking feature of life on Earth is its diversity. Consequently, the most fundamental question in ecology is which factors maintain diversity in ecological communities². Here, we analyse the combined impacts of consumers and nutrient resources on plant diversity. The supply of limiting resources, such as nutrients, controls primary productivity; that is, the rate of production of

organic matter. On local scales, productivity and diversity are often unimodally related (Fig. 1a), such that peak diversity is observed at intermediate productivity⁸. Declining diversity at higher levels of productivity is thought to be due to competitive exclusion. Exclusion can be prevented by periodic mortality events, caused by consumers or physical disturbance^{4,6,7}. These factors also show unimodal relationships with diversity (Fig. 1b). Because the effects of productivity, disturbance and consumption on diversity have been analysed separately, their interactions in nature have remained elusive. In an attempt to unify these patterns theoretically, one study explored how traditional Lotka–Volterra competition models respond to increases in productivity and disturbance frequency⁹. The study predicted that the effects of disturbance on diversity depend strongly on productivity, and vice versa (for details see Fig. 1c). Physical disturbance and consumer pressure were predicted to give similar patterns⁹. These ideas have been mathematically elaborated¹⁰, using a spatial competition model¹⁴, in which the environment consists of a large number of discrete patches, each of which can be empty or occupied by one out of n species. The model assumes a linear competitive hierarchy where species i ($1 \leq i \leq n$) would always exclude species j if $i < j$. Multi-species coexistence in this model depends on a trade-off between competitive ability and patch colonization rate c_i or extinction rate m_i (ref. 14). Productivity is assumed to enhance colonization rates of all species by a constant R and disturbance increases extinction rates of all species by a constant D (ref. 10). The dynamics of the proportion p_i of patches occupied by species i is represented as

$$\frac{dp_i}{dt} = c_i R p_i \left(1 - \sum_{k=1}^i p_k \right) - (m_i + D) p_i - \sum_{k=1}^{i-1} c_k R p_k p_i \quad (1)$$

$(i = 1, 2, \dots, n)$

where the first term represents colonization, the second local extinction and the third competitive exclusion¹⁰. Notably, predictions from this model are almost identical to those of earlier simulations^{9,10}. Thus, general patterns emerged (Fig. 1c), despite the differences in model structure (spatial compared with non-spatial), assumptions (equilibrium versus non-equilibrium) and complexity.

We tested these models in a food-web context by experimentally manipulating consumer pressure (absent, present) and nutrient supply (no, low, medium, high nutrient enrichment; see Methods)

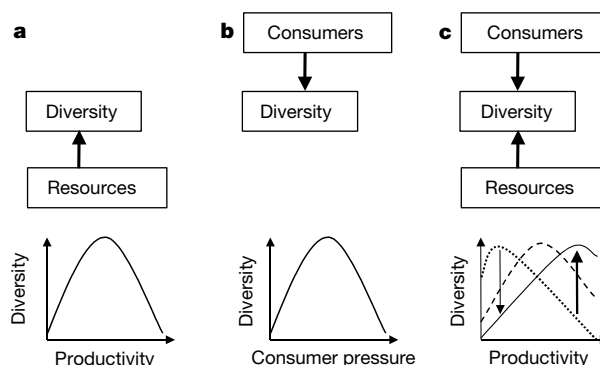


Figure 1 Consumer versus resource control of species diversity. **a**, **b**, Univariate models predict two independent relationships, where diversity peaks at intermediate resource supply or productivity (**a**), and at intermediate consumer pressure or physical disturbance (**b**), respectively^{4,5,7,8}. **c**, Multivariate models^{9,10} predict that the effects of consumers on diversity depend on resource supply and productivity; peak diversity shifts from low to intermediate to high productivity depending on whether consumer pressure is low (dotted line), intermediate (dashed line) or high (solid line). Consumers decrease diversity at low productivity (thin arrow) but increase diversity at high productivity (thick arrow).

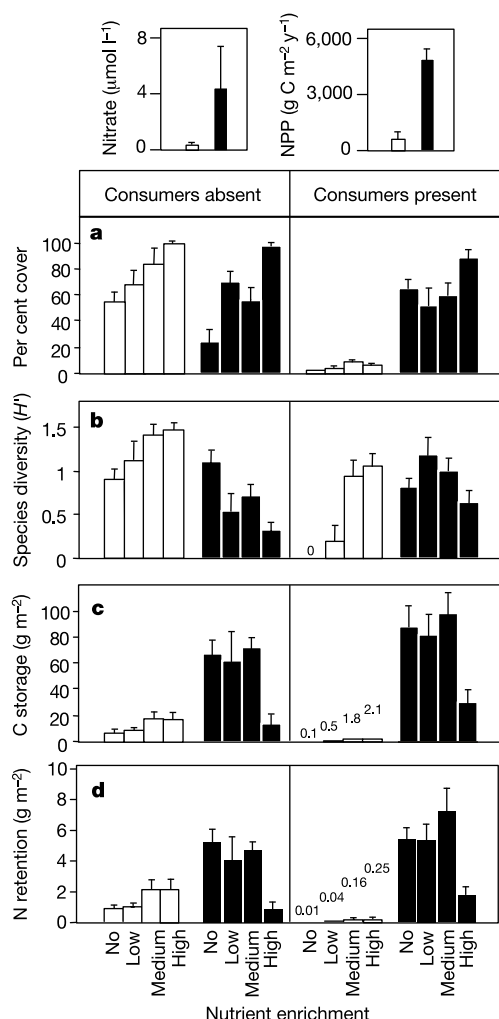


Figure 2 Factorial field experiments. Effects of consumers and nutrient enrichment on total cover (a), species diversity (b; Shannon Index (H')), carbon storage (c), and nitrogen retention (d). Open bars, Bald Rock; filled bars, Maasholm. Differences in yearly average nitrate concentrations and net primary productivity (NPP) between these sites are shown in the inserts above. All bars represent means ± 1 standard error ($n = 4$, except nitrate ($n = 10$) and NPP ($n = 8$)).

in two wave-sheltered rocky shore communities. At sheltered sites productivity is tightly linked to nutrient supply^{15,16}, and consumers are the chief source of mortality¹⁷. This is in contrast to wave-exposed shores where physical disturbance is more important¹⁷. We chose the two sites to test for consumer–nutrient interactions under

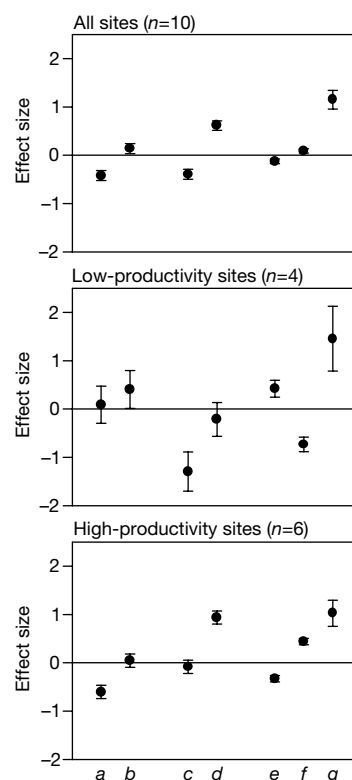


Figure 3 Factorial meta-analysis of published field experiments. Effects of nutrient enrichment on species diversity without (a) or with (b) consumers. Effects of consumers on species diversity without (c) or with (d) nutrient enrichment. Mean effects of nutrient enrichment (e), consumers (f), and consumer $\times$ nutrient interaction (g) on species diversity. Effect sizes are Hedges's d (see Methods) $\pm 95\%$ confidence interval. A positive d indicates an increase; negative d a decrease in diversity, relative to controls. Effects are statistically significant ($P < 0.05$) if confidence limits do not overlap $d = 0$.

contrasting conditions of background nutrient supply and productivity (see insert in Fig. 2). Our rationale was that to cover the full range of nutrient supply and productivity in nature we needed to combine small-scale experimental and large-scale environmental gradients. Bald Rock, Scotian shelf, northwest Atlantic, showed low nutrient supply and productivity ranged at the lower end of reported values for rocky shores¹⁵, whereas Maasholm, Baltic Sea, had very high nutrient supply and productivity was comparable to maximum values in coral reefs¹⁸.

Macroalgae were the main space colonizers in both experiments. Nutrient enrichment enhanced algal productivity, as indicated by significant increases in algal cover (Fig. 2a and Table 1). Increases in cover in nutrient-enriched plots tended to be less pronounced when

Table 1 Factorial analysis of variance

Source	d.f.	Per cent cover		Species diversity		Carbon storage		Nitrogen retention	
		<i>F</i>	<i>P</i>	<i>F</i>	<i>P</i>	<i>F</i>	<i>P</i>	<i>F</i>	<i>P</i>
Bald Rock									
Consumers	1	225.9	<0.001	66.3	<0.001	177.6	<0.001	132.8	<0.001
Nutrients	3	9.0	<0.001	22.5	<0.001	9.9	<0.001	6.9	0.002
C $\times$ N	3	2.7	0.068	5.3	0.006	0.1	0.966	1.7	0.193
Residual	24								
Maasholm									
Consumers	1	0.1	0.717	5.8	0.024	5.1	0.033	3.2	0.065
Nutrients	3	15.9	<0.001	3.9	0.021	12.7	<0.001	9.2	<0.001
C $\times$ N	3	5.1	0.007	3.0	0.051	0.5	0.714	0.5	0.797
Residual	24								

The effects of consumers, nutrient enrichment, and their interaction (C $\times$ N) on per cent cover, species diversity and ecosystem functioning in the two field experiments are shown. *F* statistics and *P*-values are shown. d.f., degrees of freedom.

consumers were present, as indicated by statistical interactions between the two factors (Table 1, note that this interaction is marginal in Bald Rock, $P = 0.068$). As predicted, changes in nutrient supply and consumer pressure had interactive effects on species diversity (Fig. 2b and Table 1). Furthermore, their main effects changed in sign between two sites of contrasting productivity (Fig. 2b). In Bald Rock, nutrient enrichment increased diversity and consumers decreased diversity. These effects were interactive: consumers had strong negative effects under ambient conditions, but weak effects under enriched conditions. In Maasholm, the reverse applied: nutrient enrichment decreased diversity and consumers increased diversity. Again, these effects were interactive: consumers reduced diversity under ambient conditions, but enhanced it under enriched conditions. Peak diversity was found in treatments without consumers in Bald Rock, but in treatments with consumers in Maasholm (Fig. 2b). Furthermore, in Maasholm peak diversity shifted to higher levels of nutrient enrichment when consumers were present.

Ecosystem functioning showed markedly similar responses to experimental manipulations as diversity. Rates of carbon storage and nitrogen retention increased with nutrient enrichment but decreased when consumers were present in Bald Rock (Fig. 2c, d and Table 1). In Maasholm these effects were reversed. Declining C storage and N retention in Maasholm was linked to the loss of long-lived species, such as perennial fucoid algae, which are replaced by fast-growing but short-lived annual algae and phytoplankton at higher levels of eutrophication¹⁹. The responses to enrichment in both experiments were highly nonlinear: in Maasholm little change in ecosystem functioning was seen in low and medium enrichment treatments, but marked decreases were seen under high nutrient loading (Fig. 2c, d). In Bald Rock, enrichment increased rates of C storage and N retention only from low to medium, but not from medium to high treatments (Fig. 2c, d). In both experiments, marked changes in ecosystem functioning coincided with a drop in diversity. Correlations between the treatment means of diversity and log-transformed ecosystem functioning revealed nonsignificant trends in Maasholm (C storage: $r = 0.76$, $P = 0.09$; N retention: $r = 0.78$, $P = 0.08$, $n = 8$) and highly significant relationships in Bald Rock (C storage: $r = 0.96$, $P = 0.01$; N retention: $r = 0.94$, $P = 0.01$, $n = 8$). Although theory suggests that this link is due to a general diversity–functioning relationship^{2,20}, this needs to be tested empirically by experiments that manipulate species diversity and composition independently.

We used factorial meta-analysis²¹ to test whether these results are consistent across widely varying communities, including marine and freshwater phytoplankton, periphyton (benthic microalgae), macroalgae and salt marshes. We compiled diversity data from ten recent field experiments that manipulated consumers and nutrient supply in factorial combination (see Methods). When we pooled the data across all sites (Fig. 3, top panel), both the effects of nutrient enrichment and consumers on diversity changed from negative to positive depending on the presence of the other factor (Fig. 3; a–d). Overall, nutrient enrichment had slight negative effects, whereas consumers had slight positive effects on diversity (Fig. 3; e, f); however, a highly significant interaction term ($P < 0.001$) suggested that it is not informative to analyse these factors in isolation (Fig. 3; g). The positive term indicates a synergistic interaction: positive effects of nutrient enrichment on diversity were only realized when consumers were present and vice versa. The Q-test statistic suggested that the effect sizes were not homogenous ($P < 0.05$). To remove this underlying heterogeneity we split the data set into low-productivity and high-productivity sites (Fig. 3, middle and bottom panels). The results show that strong interactive effects among consumers and nutrients remain (Fig. 3; g), but their mean effects on diversity changed from positive to negative (nutrients, Fig. 3; e) and from negative to positive (consumers, Fig. 3; f) with increasing productivity ($P < 0.05$).

We conclude that multivariate models^{9,10} and factorial experiments from different aquatic communities show the same striking patterns: the effects of nutrient enrichment depend on consumer pressure and vice versa. Both factors have strong and opposing effects on diversity, which change in sign among low-productivity and high-productivity ecosystems. When consumers are present, peak diversity shifts towards higher levels of nutrient supply. Moreover, our data suggest that these patterns extend to important ecosystem functions such as C storage and N retention. To our knowledge, the present study is the first where the interactive effects of consumers and resources on diversity and ecosystem functioning are shown at the same time. Three ecological theories are combined, one emphasizing resources, one emphasizing food-web interactions, and one emphasizing linkages between diversity and ecosystem functioning. This empirical and theoretical synthesis may provide a unifying concept of both causes and consequences of diversity.

These results have important implications for conservation of biodiversity and environmental management because they strongly suggest the potential for synergistic interactions among the most common human impacts on ecosystems. Human alterations of the nitrogen and phosphorus cycles continue to increase nutrient supply and productivity in terrestrial, freshwater and coastal ecosystems worldwide¹¹. At the same time, consumer pressure is altered through overharvesting of herbivore and predator populations^{12,13}, habitat fragmentation²² and destruction²³. We conclude that it is not meaningful to assess or manage these impacts in isolation. Rapid change in species composition and loss of diversity will occur when the dynamic balance of consumer and resource control is distorted, especially when consumer removals and resource enrichment occur at the same time. Our data also suggest that such changes will compromise the ability of these systems to retain the excess carbon and nitrogen that is brought upon them by human activities. □

Methods

Field experiments

Experiments were run from February to December 1998 at Maasholm, Baltic Sea (54° 41.3' N, 10° 0.5' E), and from February to December 1999 at Bald Rock, Scotian shelf, northwest Atlantic (44° 28.3' N, 63° 34.7' W) at 0.8–1 m depth below mean water level. These sites had very similar physical characteristics and species pools²⁴, but represented opposite extremes in primary productivity (see inserts in Fig. 2). Both sites were colonized by perennial fucoid macroalgae, and a larger number of fast-growing annual macroalgae (for details see ref. 24). Consumers of algal biomass were gammarid amphipods, isopods and littorinid snails at both sites. We followed algal colonization on replicate granite rocks, collected at the experimental sites. Rocks were randomly assigned to 32 cages (25 × 25 × 25 cm), covered with a clear 1-mm polyethylene mesh. Photon flux in the cages was reduced by less than 8% (LICOR SA 192-A). Consumer pressure and nutrient enrichment were manipulated using a factorial 2 × 4 design with four replicates. Consumers were excluded from closed cages but had free access to open cages, which had one side cut open. Cages were brush-cleaned weekly and checked carefully for invading consumers. Nutrient enrichment was manipulated with diffusers filled with slow-release fertilizer^{19,25}. We maintained four enrichment levels over the experimental period. These increased dissolved inorganic nitrogen by 0% (no), 8% (low), 38% (medium) and 150% (high) relative to background concentrations²⁵. Species cover was assessed using a Plexiglas sampling frame with 50 randomly placed dots. Carbon and nitrogen storage in algal biomass was analysed at the end of the growing period. All biomass was scraped off, dried at 80 °C for 48 h, and dry mass was determined to the nearest mg. Samples were ground to powder and analysed for carbon and nitrogen content on an automated C:N analyser (Fisons Instruments, NA 1500 N). Statistical analyses were performed on per cent cover data collected in late summer, when the number of species peaked at our sites. Species diversity was computed from the cover data, using the Shannon Diversity Index $H' = -\sum_{i=1}^k \ln(p_i)/p_i$, where p_i is the cover of species i divided by the total cover of k species. This index reflects species richness S (number of species), evenness ($H'/\log(S)$) and their intercorrelations, and is considered the best measure of their joint influence²⁶. Expressing diversity as richness, however, gave the same patterns (data not shown). The interactive effects of consumers and nutrient enrichment were analysed by factorial fixed-factor analysis of variance (ANOVA). Per cent cover data were angular transformed and other data were $\log(x + 1)$ -transformed to achieve homogeneity of variances.

Meta-analysis

Using a new factorial meta-analysis technique²¹ we analysed ten field experiments that reported effects of consumers and nutrient supply on species diversity. Studies had to fulfil the following criteria: consumers and nutrient supply were manipulated in a well-replicated factorial design, species diversity (Shannon Index) or species density (species

per sample area) were measured as dependent variable and treatment means, sample sizes and variance estimates were reported. Included were two experiments on macroalgae (this study), five experiments with periphyton in freshwater, brackish and marine ecosystems^{27,28}, two experiments with salt marsh plants²⁹, and one with lake phytoplankton³⁰, including subtropical and temperate climates in North America and Europe. We analysed data from sampling dates when species richness reached the seasonal peak, which was usually in late spring or summer. Data were standardized using the common meta-analysis metric of standardized effect size, Hedges's *d* (ref. 21). This is a measure of the difference between experimental and control means, divided by a pooled standard deviation and multiplied by a correction factor to account for small sample sizes. Homogeneity of effect sizes was tested using the *Q*-statistic²¹. As we detected significant heterogeneity among effect sizes we split the data set into low-productivity (oligotrophic and mesotrophic) and high-productivity (eutrophic) sites, based on information provided in the publications.

Received 23 January; accepted 12 April 2002; doi:10.1038/nature00830.

1. Hutchinson, G. E. Homage to Santa Rosalia, or why are there so many kinds of animals. *Am. Nat.* **93**, 145–159 (1959).
2. Tilman, D. Causes, consequences and ethics of biodiversity. *Nature* **405**, 208–211 (2000).
3. Sommer, U. & Worm, B. (eds) *Competition and Coexistence* (Springer, Berlin, 2002).
4. Connell, J. H. Diversity in tropical rain forests and coral reefs. *Science* **199**, 1302–1310 (1978).
5. Tilman, D. *Resource Competition and Community Structure* (Princeton Univ. Press, Princeton, 1982).
6. Flöder, S. & Sommer, U. Diversity in planktonic communities: an experimental test of the intermediate disturbance hypothesis. *Limnol. Oceanogr.* **44**, 1114–1119 (1999).
7. Buckling, A., Kassen, R., Bell, G. & Rainey, P. B. Disturbance and diversity in experimental microcosms. *Nature* **408**, 961–964 (2000).
8. Kassen, R., Buckling, A., Bell, G. & Rainey, P. B. Diversity peaks at intermediate productivity in laboratory microcosms. *Nature* **406**, 508–512 (2000).
9. Huston, M. A. *Biological Diversity* (Cambridge Univ. Press, Cambridge, 1994).
10. Kondoh, M. Unifying the relationships of species richness to productivity and disturbance. *Proc. R. Soc. Lond. B* **268**, 269–271 (2001).
11. Vitousek, P. M. *et al.* Human alteration of the global nitrogen cycle: sources and consequences. *Ecol. Appl.* **7**, 737–750 (1997).
12. Pauly, D., Christensen, V., Dalsgaard, J., Froese, R. & Torres, F. Jr Fishing down marine food webs. *Science* **279**, 860–863 (1998).
13. Jackson, J. B. C. *et al.* Historical overfishing and the recent collapse of coastal ecosystems. *Science* **293**, 629–638 (2001).
14. Tilman, D. Competition and biodiversity in spatially structured habitats. *Ecology* **75**, 2–16 (1994).
15. Mann, K. H. Seaweeds: their productivity and strategy for growth. *Science* **182**, 975–981 (1973).
16. Nielsen, K. J. Bottom-up and top-down forces in tide pools: test of a food chain model in an intertidal community. *Ecol. Monogr.* **71**, 187–217 (2001).
17. Menge, B. A. & Sutherland, J. P. Community regulation: variation in disturbance, competition, and predation in relation to environmental stress and recruitment. *Am. Nat.* **130**, 730–757 (1987).
18. Sorokin, Y. I. *Coral Reef Ecology* (Springer, Berlin, 1995).
19. Worm, B., Lotze, H. K. & Sommer, U. Coastal food web structure, carbon storage and nitrogen retention regulated by consumer pressure and nutrient loading. *Limnol. Oceanogr.* **45**, 339–349 (2000).
20. Chapin, F. S. III *et al.* Consequences of changing biodiversity. *Nature* **405**, 234–242 (2000).
21. Gurevitch, J., Morrison, J. A. & Hedges, L. V. The interaction between competition and predation: a meta-analysis of field experiments. *Am. Nat.* **155**, 435–453 (2000).
22. Terborgh, J. *et al.* Ecological meltdown in predator-free forest fragments. *Science* **294**, 1923–1925 (2001).
23. Watling, L. & Norse, E. A. Disturbance of the seabed by mobile fishing gear: a comparison to forest clearcutting. *Conserv. Biol.* **12**, 1180–1197 (1998).
24. Lotze, H. K., Worm, B. & Sommer, U. Strong bottom-up and top-down control of early life stages of macroalgae. *Limnol. Oceanogr.* **46**, 749–757 (2001).
25. Worm, B., Reusch, T. B. H. & Lotze, H. K. *In situ* nutrient enrichment: methods for marine benthic ecology. *Internat. Rev. Hydrobiol.* **85**, 359–375 (2000).
26. Stirling, G. & Wilsey, B. Empirical relationships between species richness, evenness, and proportional diversity. *Am. Nat.* **158**, 286–299 (2001).
27. Hillebrand, H., Worm, B. & Lotze, H. K. Marine microbenthic community structure regulated by nitrogen loading and herbivore pressure. *Mar. Ecol. Prog. Ser.* **204**, 27–38 (2000).
28. Hillebrand, H. & Kahlert, M. Effect of grazing and nutrient supply on periphyton biomass and nutrient stoichiometry in habitats of different productivity. *Limnol. Oceanogr.* **46**, 1881–1898 (2001).
29. Gough, L. & Grace, J. B. Herbivore effects on plant species density at varying productivity levels. *Ecology* **79**, 1586–1594 (1998).
30. Proulx, M., Pick, F. R., Mazumder, A., Hamilton, P. B. & Lean, D. R. S. Experimental evidence for interactive impacts of human activities on lake algal species richness. *Oikos* **76**, 191–195 (1996).

Acknowledgements

We thank L. Gough, R. Karez, D. Kehler, I. Milewski, R. A. Myers, R. T. Paine and T. B. H. Reusch for comments, and J. Gurevitch for statistical advice. This work was funded by the German Research Council (DFG) and the German Ministry of Science and Education.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.W. (e-mail: bworm@is.dal.ca).

A global analysis of *Caenorhabditis elegans* operons

Thomas Blumenthal*, Donald Evans*, Christopher D. Link†, Alessandro Guffanti‡, Daniel Lawson‡, Jean Thierry-Mieg§, Danielle Thierry-Mieg§, Wei Lu Chiu||, Kyle Duke¶, Moni Kiraly¶ & Stuart K. Kim¶

* Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Box B121, 4200 E. 9th Avenue, Denver, Colorado 80262, USA

† Institute of Behavioral Genetics, Box 447, University of Colorado, Boulder, Colorado 80309, USA

‡ The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

§ Gene Network Laboratory, National Institute of Genetics, Mishima 411, Japan, and National Center for Biotechnology Information, Bethesda, Maryland, USA

|| Department of Molecular Sciences and Technologies, Pfizer Global Research & Development—Ann Arbor, 2800 Plymouth Road, Ann Arbor, Michigan 48105, USA

¶ Departments of Developmental Biology and Genetics, Stanford University Medical Center, 279 Campus Drive, Stanford, California 94305, USA

The nematode worm *Caenorhabditis elegans* and its relatives are unique among animals in having operons¹. Operons are regulated multigene transcription units, in which polycistronic pre-messenger RNA (pre-mRNA coding for multiple peptides) is processed to monocistronic mRNAs. This occurs by 3' end formation and *trans*-splicing using the specialized SL2 small nuclear ribonucleoprotein particle² for downstream mRNAs¹. Previously, the correlation between downstream location in an operon and SL2 *trans*-splicing has been strong, but anecdotal³. Although only 28 operons have been reported, the complete sequence of the *C. elegans* genome reveals numerous gene clusters⁴. To determine how many of these clusters represent operons, we probed full-genome microarrays for SL2-containing mRNAs. We found significant enrichment for about 1,200 genes, including most of a group of several hundred genes represented by complementary DNAs that contain SL2 sequence. Analysis of their genomic arrangements indicates that >90% are downstream genes, falling in 790 distinct operons. Our evidence indicates that the genome contains at least 1,000 operons, 2–8 genes long, that contain about 15% of all *C. elegans* genes. Numerous examples of co-transcription of genes encoding functionally related proteins are evident. Inspection of the operon list should reveal previously unknown functional relationships.

In order to search the genome for mRNAs that contain SL2, we hybridized microarrays containing 17,817 predicted genes (94% of known and predicted genes) with probe enriched for SL2-containing mRNAs (see Methods). The results are presented in Fig. 1a. The line shows that the genes form three peaks, a peak of about 1,200 genes with very high SL2/poly(A)⁺ ratios and two larger peaks with low SL2/poly(A)⁺ ratios containing the remainder of the genes. As a positive control, we identified 319 genes that produce SL2-containing mRNAs on the basis of analysis of the sequence traces of cDNAs from the Y. Kohara laboratory (listed in Supplementary Information Table 1). Fig. 1a shows that most (84%) of these were among the SL2-enriched genes. Negative controls include 100 genes that are the first genes in the operons identified by the 100 highest SL2/poly(A)⁺ scores, and very few of these are among the SL2-enriched genes (Fig. 1b). We conclude that the microarray probing successfully identified genes that are *trans*-spliced to SL2.

Having performed a global search for genes that produce SL2 mRNAs, we determined whether their genomic structure indicated that they are located within operons. Each gene was evaluated as to whether it was likely to be downstream in an operon by the criteria described in Fig. 1 legend, using either the WormBase⁵ or the

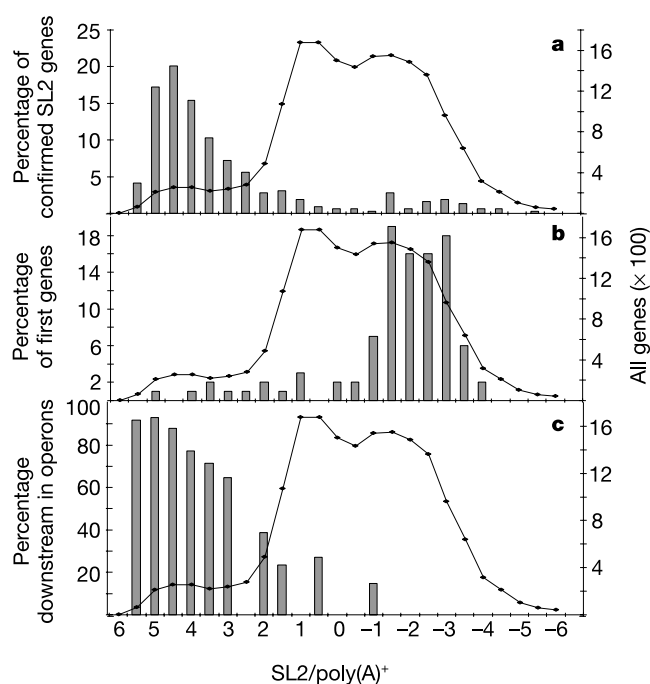


Figure 1 SL2/poly(A)⁺ ratios of 17,817 *C. elegans* genes. Genes were divided into bins according to ratios, and plotted as log₂(ratio) (line). **a**, Distribution of confirmed SL2-accepting genes. Percentage of 319 genes shown to be SL2 *trans*-spliced on the basis of sequenced cDNAs (bars). **b**, Distribution of first genes in operons. First genes in the operons identified by the 100 highest SL2/poly(A)⁺ ratios were distributed into bins. **c**, Genes in the leftmost peak and four control groups of 100 genes were evaluated for location in operons. Genes whose *trans*-splice sites were within 1 kb of the stop codon or 500 bp from the poly(A) site of another gene were scored as downstream in operons. Percentage of genes in each bin scored as downstream in operons is shown.

Intronator website⁶. In the set of 1,200 SL2-enriched genes contained in the leftmost peak, 86% were scored as downstream in operons, and only 4.5% were scored as first genes in operons (Fig. 1c). From the set of genes that do not show significant SL2/poly(A)⁺ ratios, only 15–20% were scored as possibly downstream in operons. This analysis provides strong evidence that the microarray experiment effectively identified *C. elegans* genes that are in operons. These data show a robust correlation across the genome between SL2 *trans*-splicing and downstream location in an operon, confirming and extending previous data based on individual genes.

We used three methods to estimate the number of operons in the genome. First, we collected all of the genes in operons, both from microarray data and in the list of SL2-containing cDNAs. The combined list contains 2,291 genes in 881 operons (Supplementary Information Table 2). Second, we estimated the number of operons that were missed by the microarray data. The list of SL2 spliced genes identified in the microarray experiments contained 74% of the genes identified from cDNA clones, and thus presumably of all SL2 spliced genes. Therefore we estimate that the genome contains at least 1,068 operons (790/0.74). Third, genes can be predicted to

Table 1 The number of genes per operon

Genes per operon	No. of operons
2	549
3	207
4	75
5	33
6	13
7	3
8	1

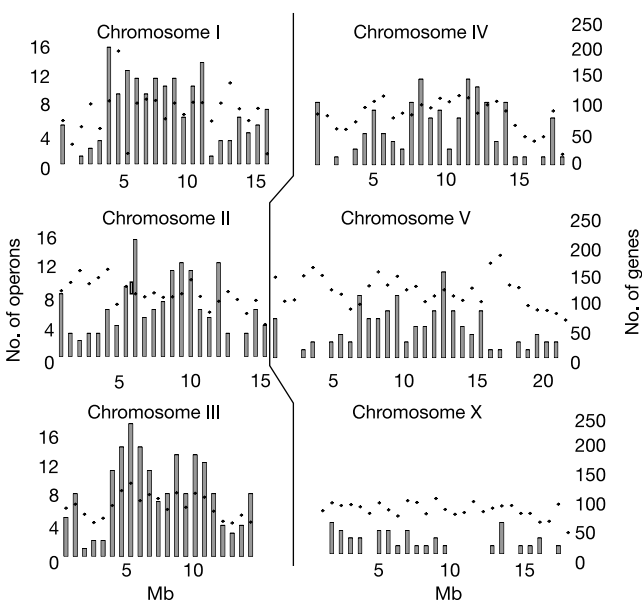


Figure 2 Chromosomal distribution of operons. Each chromosome was divided into equal-sized bins of 665,230 bp. The x-axis is in Mb from the left end of each chromosome. The number of predicted genes in each bin (right-hand y-axis) is shown by the data points. Operons (left-hand y-axis) are shown as bars.

be in operons on the basis of their gene structure. We formed a list of possible operons on the basis of gene orientation and a spacing of less than 1 kilobase (kb) between stop and start codons. There are >3,000 possible operons on this list, and 790 of these were found to be SL2-enriched in our microarray experiments. On average, the remaining genes express transcripts that are at comparable levels to the SL2-containing transcripts, making it unlikely that we missed many genes because they are expressed at too low a level to have been detected on the microarrays or by cDNA clones. Instead, the remaining genes may not be in operons, but instead may be genes that are fortuitously close together.

The average operon contains 2.6 genes, and the longest contains 8 genes (Table 1). 332 operons have more than two genes, and in 58% of these every downstream gene was scored as SL2 *trans*-spliced. These data indicate that a large percentage of SL2-accepting genes were identified, and provide strong support for the conclusion that downstream genes in operons are usually or always *trans*-spliced by SL2. If there are about 1,000 operons with 2.6 genes per operon, there are ~2,600 genes in operons. Thus the *C. elegans* genome, which contains between 17,300 (estimated from expressed open reading frames⁷) and 19,000 (all known and predicted open reading frames⁵) genes, expresses at least 13–15% of its genes as operons.

Table 2 Operons containing human disease gene orthologues

Gene	Disease	No. of genes in operon
B0261.2	Ataxia telangiectasia (ATM)	3
C01G8.5	Neurofibromatosis, type 2/Batten's disease	4
C15F1.7	Amyotrophic lateral sclerosis (ALS)	3
C16C2.3	Lowe syndrome	3
C48E7.4	Primary open angle glaucoma	4
F12F6.3	Hereditary multiple exostoses	2
F53H8.1	Hermansky–Pudlak syndrome	2
F59G1.7	Friedreich ataxia (FRDA)	7
K08E3.7	Parkinson's disease, juvenile 2	2
Y110A7A.5	Myotubular myopathy	5
Y56A3A.13	Fragile histidine triad	3
Y76A2A.2	Menkes syndrome/Wilson disease	5
ZK675.1	Nevoid basal cell carcinoma syndrome	2

An expanded version of this list, containing hypothesized functions for all of the genes in the operons, can be found in Supplementary Information.

These operons are not evenly distributed on the *C. elegans* chromosomes (Fig. 2). The X chromosome has only 37 identified operons (2.1 per megabase, Mb), whereas chromosome III has 207 (16.2 per Mb). The availability of thousands of cDNA clones allowed estimation of the distance between genes for 285 operon gene pairs (Fig. 3). The mean intergenic distance is 126-base pairs (bp), with most between 100 and 120 bp.

The correlation between SL2 *trans*-splicing and downstream position in an operon is quite strong. Nonetheless some genes that appear to be downstream in operons do not have high SL2/poly(A)⁺ scores, perhaps because their mRNAs were not well represented in the probe RNA population. Some operons that are expressed at low levels may have been missed. Also, some downstream genes in operons may get *trans*-spliced to SL1 rather than SL2⁸. Operons with long spacing might be missed because they have a tendency to be SL1 spliced³. Furthermore, some genes that do get SL2 *trans*-spliced appear not to be downstream in operons. Perhaps there is a rare mode of SL2 *trans*-splicing that does not require a gene to be downstream in an operon.

Operons are a common form of gene organization in bacteria and archaea, but they are in general absent in eukaryotes (with the possible exception of trypanosomes). Based on genome sequences of yeast, *Arabidopsis*, *Drosophila* and humans, operons are very unlikely to be found in this wide array of species. *Trans*-splicing appears to be an enabling characteristic. Presumably operons exist only when *trans*-splicing can provide a cap to protect the downstream RNA following 3' end cleavage and prevent the accompanying transcription termination. Operons have been reported only in rhabditid nematodes⁹, although recent work suggests they are found elsewhere among the nematodes (D. G. Giliano and M. Blaxter, personal communication). Nevertheless, the fact that operon organization in *C. elegans* is so common implies that the genome may be quite plastic, perhaps owing to chromosomal rearrangements producing new gene juxtapositions¹⁰. Given the relatively compact *C. elegans* genome, operon evolution may have been driven in part by constraints on chromosomal structure or organization.

Caenorhabditis elegans operons appear to be a means to co-regulate functionally related proteins, like bacterial operons. Related genes do occur in operons^{11–15}. Indeed, numerous additional examples are found in the list of operons reported here. For example, *D1054.2*, encoding a proteasome subunit, is in an operon with a ubiquitin ligase complex subunit. *ZK856.9*, which encodes a TFIIIC transcription factor, is in an operon with an RNA polymerase III subunit. *C15H11.9*, encoding a regulator of ribosome synthesis, is in an operon with an RNA polymerase I subunit. *C15C7.1*, encoding a vesicle docking and trafficking protein, is in an operon with a GRIP domain protein that also functions in the *trans*-Golgi. These and numerous other examples show that related genes

are often found together in operons. Furthermore, such relationships occur far more frequently than would be expected by chance. For example, all seven genes with an RNA-binding domain of the 'RNA recognition motif' (RRM) type that are in operons with other genes with identified functions are in operons with other nucleic-acid-interacting proteins. In contrast, of seven proteins likely to be involved with the Golgi, only one operon contains a nucleic-acid-binding protein, whereas four contain proteins related to transport. Our results show that genes for mitochondrial proteins have a strong tendency to be in operons with genes for other mitochondrial proteins, and that this relationship is highly significant ($P = 3.6 \times 10^{-4}$; see Supplementary Information Tables 3 and 4). The same is true for genes encoding splicing proteins. However, whether operons usually contain genes of related function is not yet known.

Nonetheless, the presence of a gene in an operon with another gene has recently been used to successfully predict a previously unknown functional relationship¹⁶, suggesting that the operons can be used to uncover related genes. We note that many examples of genes in operons are apparent orthologues of genes that cause disease in humans¹⁷ (Table 2). It may be possible to identify novel genes that are functionally related to the disease genes by investigating the other genes in these operons. □

Methods

SL2-enriched cDNA was prepared by reverse transcribing 5 µg of mixed stage poly(A)⁺ RNA primed with oligo(dT)¹⁸. The cDNA was denatured at 70 °C for 2 min, and annealed to a T7/SL2 primer (1 µM; 5'-TGAATTGTAATACGACTCACTATAGGAGAGGGTTTAACCCAGTTACTCA-3') at 42 °C for 5 min, followed by extension with *Escherichia coli* DNA polymerase I Klenow fragment in 100 µl at 37 °C for 30 min. RNase H was destroyed by incubating with 0.5% SDS and 20 µg proteinase K for 1 h at 55 °C. The cDNA was extracted with phenol, phenol/chloroform, chloroform/isoamyl alcohol and ethanol precipitated. SL2-enriched cRNA was prepared using T7 RNA polymerase using the manufacturer's Megascript protocol (Ambion). DNA microarrays are described in ref. 19. RNA preparation, cDNA synthesis, labelled cDNA preparation, microarray hybridization and microarray scanning were performed as previously described¹⁸. Cy3-dUTP was used to label SL2-enriched cDNA and Cy5-dUTP was used to label cDNA from poly(A)⁺ RNA made from a mixed stage population of wild-type worms. The SL2-enriched probe and the probe from the starting poly(A)⁺ mRNA were simultaneously hybridized to DNA microarrays. To ensure reproducibility, this procedure was repeated five times. Ratios of Cy3/Cy5 (SL2/poly(A)⁺) signals were calculated for each gene and converted to log₂(ratio). We then calculated the average log₂(ratio) from the five repeats. The full data set is available as Supplementary Information Table 5. The results are presented by dividing the resulting log₂(ratios) into bins (Fig. 1a).

Received 9 November 2001; accepted 15 April 2002; doi:10.1038/nature00831.

1. Spieth, J., Brooke, G., Kuersten, S., Lea, K. & Blumenthal, T. Operons in *C. elegans*: Polycistronic mRNA precursors are processed by *trans*-splicing of SL2 to downstream coding regions. *Cell* **73**, 521–532 (1993).
2. Huang, X.-Y. & Hirsh, D. A second *trans*-spliced RNA leader sequence in the nematode *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **86**, 8640–8644 (1989).
3. Blumenthal, T. & Steward, K. in *C. elegans II* (eds D. L. Riddle *et al.*) 117–145 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1997).
4. Zorio, D. A. R., Cheng, N., Blumenthal, T. & Spieth, J. Operons represent a common form of chromosomal organization in *C. elegans*. *Nature* **372**, 270–272 (1994).
5. Stein, L., Sternberg, P., Durbin, R., Thierry-Mieg, J. & Spieth, J. WormBase: network access to the genome and biology of *Caenorhabditis elegans*. *Nucleic Acids Res.* **29**, 82–86 (2001).
6. Kent, W. J. & Zahler, A. M. The intronator: exploring introns and alternative splicing in *Caenorhabditis elegans*. *Nucleic Acids Res.* **28**, 91–93 (2000).
7. Raboul, J. *et al.* Open-reading-frame sequence tags (OSTs) support the existence of at least 17,300 genes in *C. elegans*. *Nature Genet.* **27**, 332–336 (2000).
8. Williams, C., Xu, L. & Blumenthal, T. SL1 *trans*-splicing and 3' end formation in a unique class of *Caenorhabditis elegans* operon. *Mol. Cell. Biol.* **19**, 376–383 (1999).
9. Evans, D. *et al.* Operons and SL2 *trans*-splicing exist in nematodes outside the genus *Caenorhabditis*. *Proc. Natl. Acad. Sci. USA* **94**, 9751–9756 (1997).
10. Huynen, M. A., Snel, B. & Bork, P. Inversions and the dynamics of eukaryotic gene order. *Trends Genet.* **17**, 304–306 (2001).
11. Page, A. P. Cyclophilin and protein disulphide isomerase genes are co-transcribed in a functionally related manner in *Caenorhabditis elegans*. *DNA Cell Biol.* **16**, 1335–1343 (1997).
12. Huang, L. S., Tzou, P. & Sternberg, P. W. The *lin-15* locus encodes two negative regulators of *Caenorhabditis elegans* vulval development. *Mol. Biol. Cell* **5**, 395–412 (1994).
13. Clark, S. G., Lu, X. & Horvitz, H. R. The *Caenorhabditis elegans* locus *lin-15*, a negative regulator of a tyrosine kinase signalling pathway, encodes two different proteins. *Genetics* **137**, 987–997 (1994).
14. Treinin, M., Gillo, B., Liebman, L. & Chalfie, M. Two functionally dependent acetylcholine subunits are encoded in a single *Caenorhabditis elegans* operon. *Proc. Natl. Acad. Sci. USA* **95**, 15492–15495 (1998).
15. Mazroui, R., Puoti, A. & Kramer, A. Splicing factor SF1 from *Drosophila* and *Caenorhabditis*: presence of an N-terminal RS domain and requirement for viability. *RNA* **5**, 1615–1631 (1999).

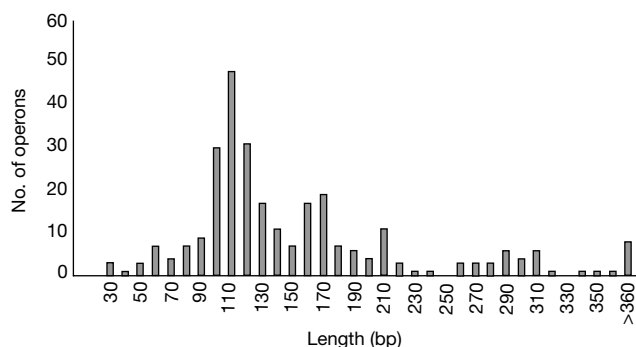


Figure 3 Operon intergenic distances. Distances from the 3' end formation site of upstream genes and *trans*-splice sites of downstream genes are plotted for the 285 operons for which reliable data are available (listed in Supplementary Information Table 6).

16. Furst, J. *et al.* ICl_{in} ion channel splice variants in *Caenorhabditis elegans*. Voltage dependence and interaction with an operon partner protein. *J. Biol. Chem.* **277**, 4435–4445 (2002).
17. Culetto, E. & Sattelle, D. B. A role for *Caenorhabditis elegans* in understanding the function and interactions of human disease genes. *Hum. Mol. Genet.* **9**, 869–877 (2000).
18. Reinke, V. *et al.* A global profile of germline gene expression in *C. elegans*. *Mol. Cell* **6**, 605–616 (2000).
19. Jiang, M. *et al.* Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **98**, 218–223 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank J. Spieth, J. Kent, A. Zahler and L. Stein for help with navigation of the *C. elegans* databases, Y. Kohara for cDNA data, M. Huang for discussions, I. Shah for statistical advice, D. Guiliano and M. Blaxter for communication of unpublished results, and P. MacMorris for advice on the manuscript. This work was supported by the NIH (T.B., C.D.L. and S.K.K.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.B. (e-mail: tom.blumenthal@uchsc.edu).

Feedback inhibition controls spike transfer in hybrid thalamic circuits

Gwendal Le Masson*, Sylvie Renaud-Le Masson†, Damien Debay‡ & Thierry Bal‡

* Laboratoire de Physiopathologie des Réseaux Neuronaux Médullaires, EPI INSERM 9914, Institut François Magendie, Université Victor Segalen Bordeaux 2, 1 Rue Camille Saint Saëns, 33077 Bordeaux Cedex, France

† Laboratoire IXL, CNRS UMR 5818, ENSEIRB, Université de Bordeaux 1, 351 Cours de la Libération, 33405 Talence Cedex, France

‡ Unité de Neurosciences Intégratives et Computationnelles, CNRS UPR 2191, Institut de Neurobiologie Alfred Fessard, 1 Avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France

Sensory information reaches the cerebral cortex through the thalamus, which differentially relays this input depending on the state of arousal^{1–5}. Such 'gating' involves inhibition of the thalamocortical relay neurons by the reticular nucleus of the thalamus^{6–8}, but the underlying mechanisms are poorly understood. We reconstructed the thalamocortical circuit as an artificial and biological hybrid network *in vitro*. With visual input simulated as retinal cell activity, we show here that when the gain in the thalamic inhibitory feedback loop is greater than a critical value, the circuit tends towards oscillations—and thus imposes a temporal decorrelation of retinal cell input and thalamic relay output. This results in the functional disconnection of the cortex from the sensory drive, a feature typical of sleep states. Conversely, low gain in the feedback inhibition and the action of noradrenaline, a known modulator of arousal^{4,9,10}, converge to increase input–output correlation in relay neurons. Combining gain control of feedback inhibition and modulation of membrane excitability thus enables thalamic circuits to finely tune the gating of spike transmission from sensory organs to the cortex.

The thalamus is the major gateway for the flow of sensory information to the cerebral cortex. Far from being a passive relay, this structure actively processes information before cortical integration. It is the first stage at which sensory signals can be gated during selective attention or during the transition from general

arousal to sleep^{1–5,8}. Although much is known about the anatomy and the synaptic and cellular properties of the thalamic networks, the nature of the sensory information processing throughout selective arousal and sleep–wake stages is not yet understood. The goal of this work was to investigate the mechanisms responsible for changes in the efficiency of sensory spike transfer in the retino-thalamic network, during different states of arousal. We used hybrid biological–neuromimetic networks that allow direct control of cellular and synaptic components¹¹. We measured the variations of spike-to-spike correlation between identified input and output neurons, reflecting the efficiency and reliability of signal transfer in different activity states.

In our hybrid networks (Fig. 1a), synaptic-like interactions between realistic conductance-based model neurons and an intracellularly recorded biological neuron run in real time, following the natural dynamics of the biological cell or network. Individual membrane currents of the simulated and biological neurons and the properties of their synaptic connections can be selectively and quantitatively controlled throughout their dynamic range, *in vitro* or *in vivo*. The required speed of real-time computation is achieved by using both programmable digital signal processors (DSPs) and newly designed analog integrated circuits^{11,12}. A dynamic clamp procedure was used to simulate synaptic conductances by current injection through the intracellular recording pipette¹³.

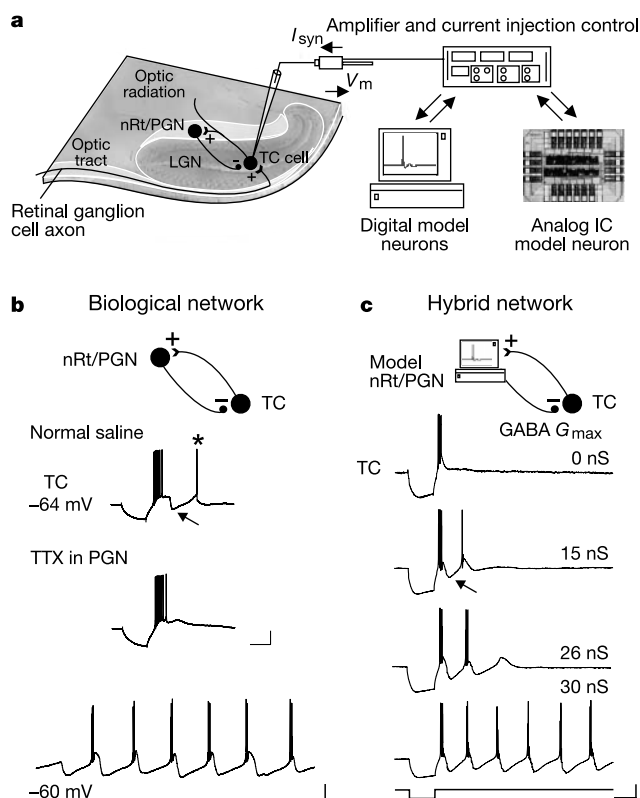


Figure 1 Design of hybrid thalamic circuits. **a**, Artificial synaptic connections between a biological TC cell recorded intracellularly in an LGNd slice and DSP-based and analog integrated circuit (IC) neurons. Wiring diagram in a ferret LGNd slice: +, excitatory; –, inhibitory. **b**, One-to-one coupling in ferret networks: a burst of spikes evoked in a single TC neuron can trigger burst firing of a target PGN neuron (not shown), which generates feedback inhibition^{14–16} (arrow) and rebound burst (asterisk). Middle, tetrodotoxin (TTX) block of PGN activity prevents feedback inhibition. Bottom, synaptic interaction between PGN and TC neurons leads to repetitive TC bursts. **c**, Hybrid circuit reconstruction using nRt/PGN model cell, in guinea-pig LGNd slices where TC cells are initially synaptically isolated: effect of incrementing nRt/PGN-mediated GABA conductance. Calibration bars, 0.1 s, 20 mV, 0.35 nA.

GABA (γ -aminobutyric acid)-releasing reticular interneurons (nucleus reticularis (nRt)/perigeniculate (PGN)) provided a strategically located inhibitory feedback to the corticopetal afferent sensory pathway (Fig. 1a), activated by recurrent collaterals of the thalamocortical (TC) cells and by corticothalamic projections^{1,7} and modulated by brainstem afferents^{1,9}. We hypothesized that changes in the gain of reticular-mediated inhibition¹⁴ affect the temporal correlation between input and output spikes, thus providing a mechanism for spike transfer control. We tested this hypothesis using canonical hybrid thalamic circuits of the lateral geniculate nucleus (LGN dorsal, LGNd). These consist of a biological TC neuron reciprocally connected to a model reticular interneuron. The marked similarity of the behaviour of the hybrid circuits compared with the known reciprocal interactions of the wholly biological TC–nRt/PGN circuit^{14–16} validates the hybrid equivalent reconstruction (Fig. 1b, c). In the wholly biological circuit, synaptic interactions between nRt/PGN and TC cells can generate sustained network oscillations, known as spindle waves⁵, a landmark of early stages of sleep. The basic properties of this intrathalamic loop (Fig. 1b) can be reproduced in the hybrid circuit with full control of the ionic conductances modulating synaptic strength (Fig. 1c). Progressive increase of the maximal conductance ($G_{\max}$) of the nRt/PGN-to-TC GABA-mediated synapse (mixed GABA_A/GABA_B conductances) from 0 to 73 nS ($n = 14$) increased the probability of rebound burst generation in the TC neuron (Fig. 1c). Oscillations resembling spindle waves (Fig. 1b, c) were obtained for a threshold value of GABA $G_{\max}$ that showed little variation between TC cells (29 ± 4.2 nS, mean $\pm$ s.e.m.; $n = 9$) (see Methods).

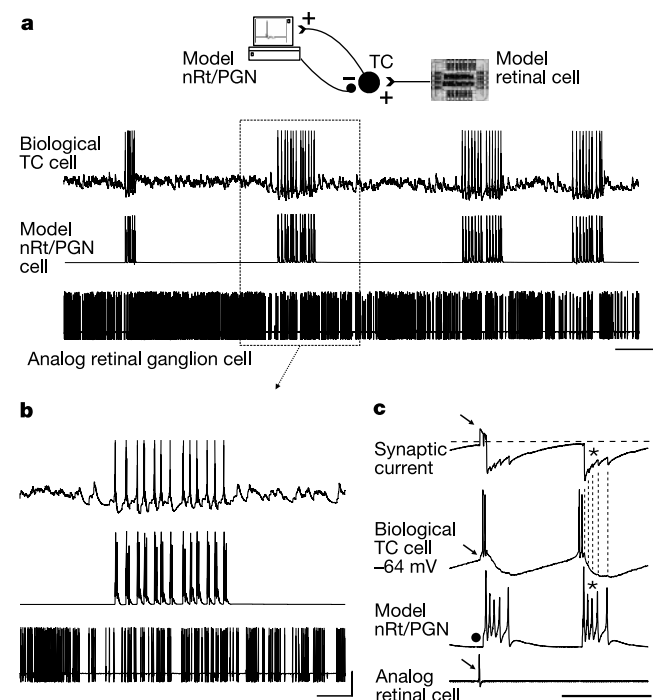


Figure 2 Spontaneous spindle activity in 'sleeping' hybrid retinothalamic circuit. **a**, Patterned firing (mean rate 42 Hz; $\gamma = 1.5$) of the analog retinal neuron generates an artificial synaptic bombardment in the biological TC cell. **b**, Detail of **a**. **c**, Total synaptic current injected into the biological TC cell and simultaneous voltage records. Outward synaptic current is downward. Horizontal dashed line indicates null net current. A positive AMPA current triggered by a retinal cell spike induced an EPSP in the biological TC cell (arrows). TC cell spikes triggered AMPA EPSPs and bursts of spikes in the nRt/PGN cell (circle). The latter triggered mixed GABA_A/GABA_B negative currents (asterisk) that summed in the TC cell. Calibration bars: 1 s, 20 mV (**a**); 0.3 s, 20 mV (**b**); 0.1 s, 20 mV, 1 nA (**c**).

We explored the signal transfer capability of thalamic loops by adding a simulated retinal input to the hybrid circuit. *In vivo*, individual TC neurons receive the bulk of their excitation from a single retinal ganglion cell¹⁷. In darkness or under constant illumination, different types of ganglion cells discharge irregularly with a mean rate ranging from 5 to 60 Hz and interspike interval (ISI) distributions adequately modelled¹⁸ as renewal processes with γ -distributed intervals (see Methods). In hybrid circuits, analog model neurons were programmed to simulate the temporal structure of retinal cell discharges and the kinetics of real AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptor-mediated excitatory postsynaptic potentials (EPSPs). Hybrid thalamic circuits receiving this realistically emulated synaptic bombardment spontaneously generated short epochs of oscillation (frequency 9.26 ± 0.87 Hz; duration 1.74 ± 0.36 s; $n = 27$), recurring periodically in a manner very similar to biological spindle waves occurring in sleep-related states (Fig. 2). Each spindling epoch is terminated by an after-depolarization that is similar in the biological and the hybrid network. This after-depolarization, resulting from the hyperpolarization-activated current I_h (ref. 19),

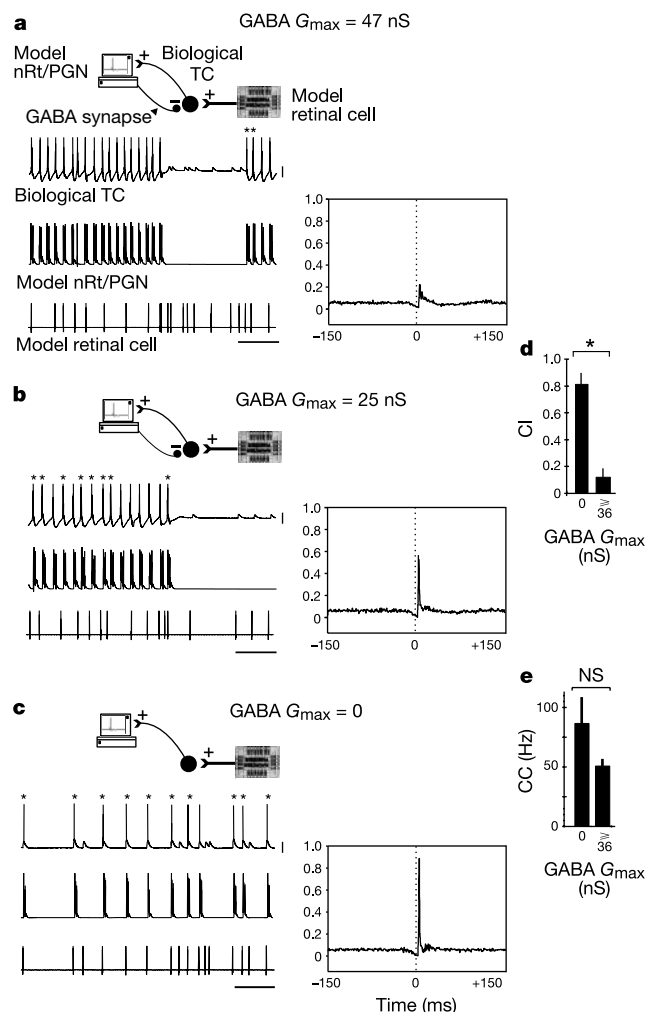


Figure 3 The strength of inhibition regulates temporal correlation between input retinal and output TC cell spikes. **a**, High-conductance, mixed GABA_A/GABA_B (GABA $G_{\max}$), nRt/PGN-to-TC inhibitory synapse. Right panels, contribution index (2-ms bins). Asterisk, TC spikes triggered within a delay of <10 ms after retinal spike (5 Hz; $\gamma = 3$). **b**, **c**, Decreasing the inhibition increases the probability of input–output correlation. **d**, **e**, Averaged contribution (CI, **d**) and correlation indices (CC, **e**) versus synaptic strength. Six screening experiments, 2-ms bins, input rate 21.6 ± 6.9 Hz, $\gamma = 3$. Mann–Whitney U -test, $P < 0.05$. Calibration bars, 0.5 s, 20 mV.

sums to the synaptic bombardment to form a depolarization that causes inactivation of the low-threshold Ca^{2+} current and subsequent block of input signal transfer.

During this emulated sleep-like state, the firing pattern of TC cells, which is very different from the retinal cell input, suggests that the transfer of retinal cell spikes to the cortex is low (Fig. 2a). The question of whether this low transfer could nonetheless be reliable was examined by calculating two different indices. First, the contribution index (CI) examines the TC cell discharge and quantifies the percentage of spikes that are precisely correlated with retinal input spikes. It estimates the reliability with which a TC spike can be considered as being triggered by the input rather than being spontaneous. It is computed as the peak of the cross-correlation normalized by the number of TC spikes (output). Second, the correlation index (CC) measures the global efficiency of the input–output spike transfer and indicates the ratio of input spikes being actually transmitted as output spikes in the TC neuron. It is computed as the peak of the cross-correlation between retinal and

TC neuron spikes, normalized by the number of retinal cell spikes (input).

In the presence of strong inhibitory feedback ($\text{GABA } G_{\max} = 47 \text{ nS}$), the CI was low (Fig. 3a), indicating that most of the TC spikes were not correlated with the retinal input and thus that the thalamus was not transferring spikes in a one-to-one manner. We then systematically screened the strength of inhibition to test whether the degree of inhibition produced by nRt/PGN interneurons could directly control the precision of the input–output temporal correlation. As the inhibitory synaptic strength ($G_{\max}$) decreased, the spike-to-spike correlation gradually increased (Fig. 3a–d). Thus, the recurrent nRt/PGN feedback inhibition has a direct de-correlating effect capable of reducing the reliability of spike transfer.

Despite these highly significant inhibition-dependent variations of CI (Fig. 3d), the non-selective global efficiency of spike transfer, measured by the CC, remained low and was not significantly different in the presence or absence of feedback inhibition (Fig. 3e). This low transfer efficiency was predicted by computational models²⁰ and can be explained by the low-threshold calcium current that, at hyperpolarized membrane potentials, endows TC cells with low-pass filtering properties typical of sleep states^{4,5}.

These results show that high gain of feedback synaptic inhibition strengthens the filtering properties of TC cells and ensures a nearly complete decorrelation of the output activity from the sensory input. The resulting combination of low reliability and low efficiency of spike transfer explains how sensory input may be disconnected within the thalamic circuit during sleep.

In contrast, during arousal, reliable synaptic transmission and short-latency responses to afferent stimuli are required^{4,5}. It is well known that membrane properties of thalamic cells are modulated by neurotransmitters released by cortical, brainstem and hypothalamic afferents in the thalamus^{9,10}, switching their activity from bursting mode during sleep-related states to an activity dominated by tonic firing during circadian or phasic episodes of arousal^{14,5,9}. The transition from sleep to waking can be mimicked *in vitro* by local application of noradrenaline²¹, known to increase the excitability of relay TC neurons and to decrease the gain of intrathalamic inhibition²² through activation of postsynaptic $\alpha 1$ - and β -adrenoceptors^{4,5,9,22}.

Here, we used hybrid circuits to investigate how the modulation of membrane properties alters the balance of cell–cell interactions, and thus also affects input–output spike correlation (Fig. 4). Noradrenaline increased both spike transfer efficiency and reliability, measured respectively by CC and CI (Fig. 4b–e; compare left with middle and right cross-correlation histograms in Fig. 4b, and bars in Fig. 4d and e), even when associated with a high value of feedback inhibition from nRt cells. Furthermore, TC spike latency became shorter and less variable (control, $5.3 \pm 1.3 \text{ ms}$, versus noradrenaline, $2.8 \pm 0.3 \text{ ms}$, $P < 0.05$). These effects were often associated with a tonic depolarization of the TC cell (Fig. 4a). In this neurotransmitter-dependent activated state, spike transfer efficiency (CC) could be further enhanced by removing the nRt/PGN-mediated inhibition (Fig. 4b–d; compare middle and right cross-correlation histograms in Fig. 4c and bars in Fig. 4d) and in four cells this increased the reliability estimate (CI; Fig. 4e). In the presence of noradrenaline, the remaining two cells tended to generate supplementary uncorrelated spikes. When synaptic inhibition was absent, this resulted in an input/output spike ratio of greater than one, which decreased CI. This suggests that the presence of feedback inhibitory postsynaptic potentials (IPSPs) in these cells can make transfer more reliable by filtering out spikes not correlated with the simulated retinal input.

Our results show that hybrid thalamic circuits sample and filter input spikes with a temporal precision that is regulated by both the strength of intrathalamic inhibitory synapses and the state of membrane properties of TC neurons (Fig. 4f). It is probable that *in vivo* these combined mechanisms constitute a versatile control

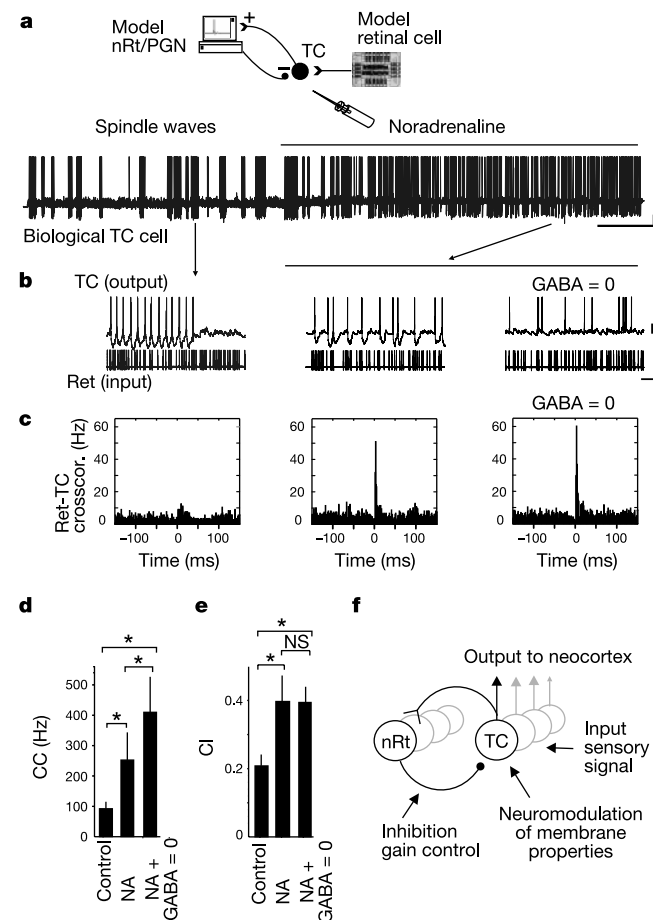


Figure 4 Noradrenaline (NA) enhances input–output correlation in TC neurons. **a**, Application of NA 500 μM (horizontal bar) on biological TC. Retinal input, 30 Hz; $\gamma = 3$. **b**, Detail from **a**: left, sleep-like state, $\text{GABA } G_{\max} = 36 \text{ nS}$; middle, $\text{GABA } G_{\max} = 36 \text{ nS}$; right, under NA, $\text{GABA } G_{\max} = 0$. Calibration bars: 10 s (**a**) or 200 ms (**b**), 20 mV. **c**, Normalized cross-correlation histograms (1-ms bins) for conditions in **b**. Left, low correlation; middle, increased correlation under NA; right, NA and no nRt/PGN inhibition. **d**, **e**, Transfer efficiency (CC, **d**) and contribution indices (CI, **e**) versus experimental conditions (averaged peaks of cross-correlation histograms, 1-ms bins, six cells). Asterisk, $P < 0.05$, nonparametric Wilcoxon matched-pairs test. $\gamma = 3$; 20 Hz for five retinal cells, 30 Hz for one cell; $\text{GABA } G_{\max} = 38 \pm 1 \text{ nS}$; $\text{AMPA } G_{\max} = 21 \pm 2 \text{ nS}$. **f**, Converging influences of the strength of feedback inhibition and membrane property changes of TC cells.

system that allows precise up- and downregulation of signal transfer, over a continuum from sleep to arousal.

These results were obtained in a reduced retinthalamic circuit, but generate principles that are applicable to the larger-scale thalamic network. Filtering of the global retinal input can be seen as a result of the interactive operation of similar canonical circuits reiterated across the retinotopic space. Functional and anatomical studies describe significant overlap between parallel sensory channels through convergent and divergent synaptic connections both between the thalamic cell types¹⁴, and within corticothalamic feedback projections. Crosstalk between channels is a functional variable, permitting the synchronization of the activity of large populations of thalamic cells during slow-wave-sleep^{4,5,23} or epileptic absence seizures^{4,24,25}. We therefore propose that during these states of low perceptual awareness, input-output spiking decorrelation caused by nRt/PGN-mediated inhibition could be generalized across the entire thalamic network, thus contributing to reduced sensory perception.

In the transition from sleep to arousal, activation of TC and nRt/PGN cells by extrinsic neuromodulatory influences decreases their ability to fire in burst mode^{5,9,22}. This, in turn, weakens feedback inhibition of TC cells^{14,26} and ultimately blocks slow recurrent circuit oscillations²¹. We postulate that, in arousal, decreased inhibition across the network results in a generalized increase in input-output spike correlation in many parallel sensory channels. This would tend to shift the time constant of integration of sensory input to the millisecond range. Such a change in network behaviour might also act in a more restricted local manner, contributing to focal attention⁸ processes. The gain of nRt/PGN-mediated inhibition could also be tuned down in restricted parts of the thalamic network by corticothalamic or brainstem afferents. This would provide a mechanism for selective sensory attention by creating islands of highly efficient sensory spike transfer in a network otherwise decoupled from its inputs. □

Methods

Biological preparations

Biological neurons were recorded at 34.5–35.5 °C from slices²⁵ of guinea-pig (1–4 months old) or ferret (2–15 months old) LGNd. Intracellular records were made with glass micropipettes filled with 1.2 M potassium acetate (90–100 MΩ after bevelling²⁵). Drugs were applied locally by the pressure-pulse technique²⁵. Hybrid networks were built in guinea-pig preparations in which TC neurons were synaptically isolated from the extrinsic afferent inputs (including nRt) in the process of making the LGNd slice. This was translated by the complete absence of spindle waves and conferred the important advantage that it was not necessary to use pharmacological blockers, which might also have affected the response properties of the cells. Mean input resistance of guinea-pig TC cells was 68.1 ± 3.1 MΩ ($n = 19$). Drifts in membrane potential were corrected, if necessary, to maintain the threshold for action potentials measured during rebound bursts at -44.4 ± 0.6 mV ($n = 11$).

Spindle generation in biological and hybrid networks

Spindle waves, generated in thalamic circuits, appeared in the electroencephalogram as 7–14-Hz oscillations that waxed and waned in amplitude over a 2–4-s period and reappeared once every 3–10 s^{4,5}. Knowledge of their mechanisms, acquired in the ferret LGNd slice^{14–16}, was used to design and validate our hybrid circuits. GABA-mediated PGN neurons generated direct inhibitory feedback to TC neurons, which resulted in intermittent activation of low-threshold calcium spikes and associated bursts of action potentials. In turn, as in the hybrid network, rebound bursts of TC cells reactivated the TC–PGN–TC feedback loop, thus entraining the circuit in spindle wave oscillation. In the hybrid circuit, rhythmic IPSPs activated a rebound burst in the biological cell at each cycle of spindle oscillation. Similar rebound bursting can be observed in the purely biological circuit (Fig. 1b, bottom trace).

Implementation of model neurons and synapses

Hybrid interaction between biological neurons and models¹¹ was based on a dynamic clamp¹³ procedure. The principal advantage of the hybrid network was the ability to selectively and quantitatively control membrane properties and synaptic gain (Fig. 1c), throughout their dynamic range, in a systematic manner. Real-time connections require fast computation, which was achieved with DSP boards, coupled with digital-to-analog and analog-to-digital (8 bits, 128 kHz) conversion stages (Innovative Integration) and controlled by dedicated software. Numerical models of thalamic neurons were single-compartment conductance-based models^{27–29}. Analog silicon integrated circuits, performing real-time processing independently of the model's complexity, were developed

using a full custom current-mode design with 1.2 μm BiCMOS technology¹². These multipurpose analog neurons were conductance-based realistic models implemented on programmable, application-specific integrated circuits (ASICs). Synaptic input from a model cell to a biological neuron is achieved by generating a voltage signal proportional to a synaptic current according to the equation $I_{\text{syn}} = G_{\text{syn}}(t, V_m)(V_r - E_s)$. This voltage signal was converted to current in an Axoclamp 2B (Axon Instruments) amplifier used in discontinuous current-clamp mode and fed to the recorded neuron via the intracellular microelectrode. G_{syn} is the synaptic conductance expressed as a function of time and presynaptic model membrane potential (V_m); unless stated in the text, in control conditions G_{syn} values were (in nS): GABA_A G_{max} 35; GABA_B G_{max} 1.4; TC to nRt AMPA G_{max} 20; and retino-TC AMPA G_{max} 28. The formalism and parameters describing the dynamics of G_{syn} for AMPA and GABA synapses were based on previous models^{28,29} that include a kinetic scheme of neurotransmitter release. V_r is the biological postsynaptic neuron membrane potential. E_s is the reversal potential for the ion species involved and was set at (in mV): Na⁺, +55; Cl[−], −90; and K⁺, −110. To reproduce natural intraspindle frequency and duration, the parameter space of GABA receptor subtypes (GABA_A G_{max} versus GABA_B G_{max}) was screened systematically and given an optimal conductance ratio of 96% GABA_A and 4% GABA_B ($n = 18$), consistent with previous findings^{4,5,14–16,25,26}.

The AMPA-type excitatory synaptic inputs from retinal ganglion cells to TC, as well as from TC to nRt, were simulated using the model of Destexhe *et al.*²⁸. Because inputs from retinal ganglion cells are located near the soma¹, on the proximal dendrites of TC cells, artificial somatic injection using dynamic clamp is a realistic procedure for mimicking retinal inputs. Simulated recording at a proximal dendritic site 35 μm from the soma of a 200-compartment TC model cell, shows that AMPA receptor-mediated EPSPs generated in the soma or the dendrite give almost identical responses (A. Destexhe, unpublished simulation). Thus, artificial AMPA EPSPs efficiently trigger low-threshold calcium spikes because the distribution of T (transient) Ca²⁺ channels is non-uniform and highest in proximal dendrites³⁰. The ISI distribution of retinal ganglion cells can be accurately described by an Erlang renewal process, where the width of the distribution is set by the γ order¹⁸. A high γ -value leads to highly regular ISI, whereas lower γ -values generate irregular ISIs. Renewal processes with γ orders within the physiological range ($\gamma = 0.7$ to 12)¹⁸ were used to generate ISIs for our analog retinal cells.

Received 3 December 2001; accepted 8 April 2002; doi:10.1038/nature00825.

- Sherman, S. M. & Guillery, R. W. Functional organization of thalamocortical relays. *J. Neurophysiol.* **76**, 1367–1395 (1996).
- Coenen, A. M. L. & Vendrick, A. J. H. Determination of the transfer ratio of cat's geniculate neurons through quasi-intracellular recordings and the relation with the level of alertness. *Exp. Brain Res.* **14**, 227–242 (1972).
- Livingstone, M. S. & Hubel, D. H. Effects of sleep and arousal on the processing of visual information in the cat. *Nature* **291**, 554–561 (1981).
- Steriade, M., Jones, E. G. & McCormick, D. A. *Thalamus, Organization and Function* 533–685 (Elsevier Science, Oxford, 1997).
- Steriade, M., McCormick, D. A. & Sejnowski, T. J. Thalamocortical oscillations in the sleeping and aroused brain. *Science* **262**, 679–685 (1993).
- Yingling, C. D. & Skinner, J. E. Gating of thalamic input to cerebral cortex by nucleus reticularis thalami. Attention, voluntary contraction and event-related cerebral potentials. *Prog. Clin. Neurophysiol.* **1**, 70–96 (1977).
- Ahlens, G., Lindström, S. & Lo, F. S. Interaction between inhibitory pathways to principal cells in the lateral geniculate nucleus of the cat. *Exp. Brain Res.* **58**, 134–143 (1985).
- Montero, V. M. Amblyopia decreases activation of the corticogeniculate pathway and visual thalamic reticularis in attentive rats: a 'focal attention' hypothesis. *Neuroscience* **91**, 805–817 (1999).
- McCormick, D. A. Neurotransmitter actions in the thalamus and cerebral cortex and their role in neuromodulation of thalamocortical activity. *Prog. Neurobiol.* **39**, 337–388 (1992).
- Aston-Jones, G., Chiang, C. & Alexinsky, T. Discharge of noradrenergic locus coeruleus neurons in behaving rats and monkeys suggests a role in vigilance. *Prog. Brain Res.* **88**, 501–520 (1991).
- Le Masson, G., Le Masson, S. & Moulins, M. From conductances to neural network properties: analysis of simple circuits using the hybrid network method. *Prog. Biophys. Mol. Biol.* **64**, 201–220 (1995).
- Le Masson, S., Laflaquière, A., Bal, T. & Le Masson, G. Analog circuits for modeling biological neural networks: design and applications. *IEEE Trans. Biomed. Eng.* **64**, 638–645 (1999).
- Sharp, A. A., O'Neil, M. B., Abbott, L. F. & Marder, E. Dynamic clamp: computer generated conductances in real neurons. *J. Neurophysiol.* **69**, 992–995 (1993).
- Kim, U. & McCormick, D. A. The functional influence of burst and tonic firing mode on synaptic interactions in the thalamus. *J. Neurosci.* **18**, 9500–9516 (1998).
- McCormick, D. A. & Bal, T. Sleep and arousal: thalamocortical mechanisms. *Annu. Rev. Neurosci.* **20**, 185–215 (1997).
- von Krosigk, M., Bal, T. & McCormick, D. A. Cellular mechanisms of a synchronized oscillation in the thalamus. *Science* **261**, 361–364 (1993).
- Usrey, W. M. & Reid, R. C. Synchronous activity in the visual system. *Annu. Rev. Physiol.* **61**, 435–456 (1999).
- Troy, J. B. & Robson, J. G. Steady discharges of X and Y retinal ganglion cells of cat under photopic illuminance. *Vis. Neurosci.* **9**, 535–553 (1992).
- Bal, T. & McCormick, D. A. What stops synchronized thalamocortical oscillations? *Neuron* **17**, 297–308 (1996).
- Mukherjee, P. & Kaplan, E. Dynamics of neurons in the cat lateral geniculate nucleus: *in vivo* electrophysiology and computational modeling. *J. Neurophysiol.* **74**, 1222–1242 (1995).
- Lee, K. H. & McCormick, D. A. Abolition of spindle oscillations by serotonin and norepinephrine in the ferret lateral geniculate and perigeniculate nuclei *in vitro*. *Neuron* **17**, 309–321 (1996).
- Funke, K., Pape, H. C. & Eysel, U. T. Noradrenergic modulation of retinogeniculate transmission in the cat. *J. Physiol. (Lond.)* **463**, 169–191 (1993).
- Contreras, D., Destexhe, A., Sejnowski, T. J. & Steriade, M. Control of spatiotemporal coherence of a thalamic oscillation by corticothalamic feedback. *Science* **274**, 771–774 (1996).
- Snead, O. C. III Basic mechanisms of generalized absence seizures. *Ann. Neurol.* **37**, 146–157 (1995).

25. Bal, T., Debay, D. & Destexhe, A. Cortical feedback controls the frequency and synchrony of oscillations in the visual thalamus. *J. Neurosci.* **20**, 7478–7488 (2000).
26. Kim, U., Sanchez-Vives, M. V. & McCormick, D. A. Functional dynamics of GABAergic inhibition in the thalamus. *Science* **278**, 130–134 (1997).
27. McCormick, D. A. & Huguenard, J. R. A model of the electrophysiological properties of thalamocortical relay neurons. *J. Neurophysiol.* **68**, 1384–1400 (1992).
28. Destexhe, A., Mainen, Z. F. & Sejnowski, T. J. Synthesis of models for excitable membranes, synaptic transmission and neuromodulation using a common kinetic formalism. *J. Comput. Neurosci.* **3**, 195–230 (1994).
29. Destexhe, A., Bal, T., McCormick, D. A. & Sejnowski, T. J. Ionic mechanisms underlying synchronized oscillations and propagating waves in a model of ferret thalamic slices. *J. Neurophysiol.* **76**, 2049–2070 (1996).
30. Williams, S. R. & Stuart, G. J. Action potential backpropagation and somato-dendritic distribution of ion channels in thalamocortical neurons. *J. Neurosci.* **15**, 1307–1317 (2000).

Acknowledgements

We are grateful to K. Grant, Y. Fregnac, S. Olet, F. Nagy, A. Destexhe, M. Rudolph and B. Gutkin for in-depth discussion and comments on the manuscript; G. Sadoc, N. Gazère and E. Barbe for their technical input; and A. Destexhe for theoretical simulations. This research was supported by the Groupement d'Intérêt Scientifique Sciences de la Cognition, Centre National de la Recherche Scientifique, Institut National de la Santé et de la Recherche Médicale, Plan Pluriformation du Ministère de la Recherche, Fondation pour la Recherche sur l'Epilepsie, and the Institut Electricité Santé de France.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.B.
(e-mail: thierry.bal@iaf.cnrs-gif.fr).

Na⁺/H⁺ exchanger regulatory factor 2 directs parathyroid hormone 1 receptor signalling

Matthew J. Mahon*, Mark Donowitz†, C. Chris Yun† & Gino V. Segre*

* Endocrine Unit, Massachusetts General Hospital and Department of Medicine, Harvard Medical School, Boston, Massachusetts 02114, USA

† Department of Medicine, Gastroenterology Division, The Johns Hopkins University School of Medicine, Boston, Massachusetts 21205, USA

The parathyroid hormone 1 receptor (PTH1R) is a class II G-protein-coupled receptor¹. PTH1R agonists include both PTH, a hormone that regulates blood calcium and phosphate, and PTH-related protein (PTHrP), a paracrine/autocrine factor that is essential for development, particularly of the skeleton. Adenylyl cyclase activation is thought to be responsible for most cellular responses to PTH and PTHrP, although many actions appear to be independent of adenylyl cyclase^{1–5}. Here we show that the PTH1R binds to Na⁺/H⁺ exchanger regulatory factors (NHERF) 1 and 2 through a PDZ-domain interaction *in vitro* and in PTH target cells. NHERF2 simultaneously binds phospholipase Cβ1 and an atypical, carboxyl-terminal PDZ consensus motif, ETVM, of the PTH1R through PDZ1 and PDZ2, respectively. PTH treatment of cells that express the NHERF2–PTH1R complex markedly activates phospholipase Cβ and inhibits adenylyl cyclase through stimulation of inhibitory G proteins (G_{i/o} proteins). NHERF-mediated assembly of PTH1R and phospholipase Cβ is a unique mechanism to regulate PTH signalling in cells and membranes of polarized cells that express NHERF, which may account for many tissue- and cell-specific actions of PTH/PTHrP and may also be relevant to signalling by many G-protein-coupled receptors.

In most cells, the activated PTH1R increases cyclic AMP robustly through stimulation of adenylyl cyclase, and increases hydrolysis of inositol phosphates only modestly through stimulation of phospholipase C (PLC)^{1,6}. The PTH1R couples to stimulatory G proteins (G_s) and protein of the G_q subfamily, but also couples to G_{i/o} protein(s), as shown by pertussis-toxin blockade of PTH activation of PLC and increased activation of adenylyl cyclase^{7–10}. In COS7 (African green monkey kidney) cells, pertussis-toxin sensitivity was observed only with intact PTH1R, but not with PTH1R, which had a carboxy-terminal deletion of 111 amino acids¹¹. In addition, PTH increased cAMP accumulation in cells expressing the truncated PTH1R by 4–6-fold that of cells expressing full-length PTH1R, but PTH-stimulated increases of inositol phosphates were unaffected. These findings suggest that the C-terminal tail contains determinants that modulate signalling by the PTH1R.

Using the C-terminal, cytoplasmic tail of the PTH1R as bait in a yeast two-hybrid screen of a human kidney complementary DNA library, we found a strong interaction with a NHERF2 fragment, termed 12A. 12A includes PDZ2 and the C-terminal ERM-binding region, but lacks amino-terminal sequences, including PDZ1. It contains a consensus splice sequence and 105 extra-exonic bases, which are spliced in-frame and encode 35 amino acids N-terminal to amino acid 139 of NHERF2. 12A is most probably an incompletely processed messenger RNA, because the 5' region could not be identified by 5' rapid amplification of cDNA ends (RACE) using a 3' primer that included a portion of 12A. The high homology between 12A and NHERF2, however, led us to study NHERF2–PTH1R interactions.

NHERF2 contains two PDZ (PSD95, discs large protein, ZO1) domains that assemble target proteins by binding to a C-terminal motif having a consensus sequence—(D/E)-(S/T)-X-(L/V)¹²—and a C-terminal ERM-binding domain that mediates interactions with the actin-binding proteins ezrin, radixin and moesin¹³. NHERF2

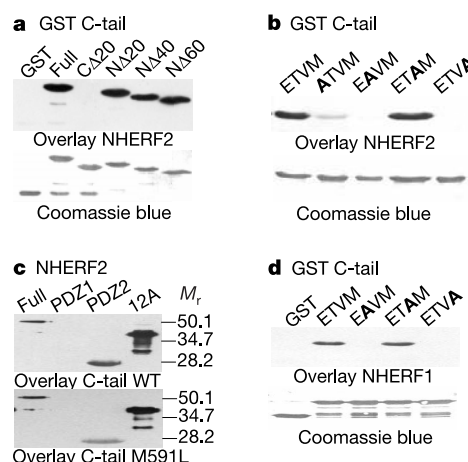


Figure 1 *In vitro* interactions of PTH1R and NHERF1 and 2 are PDZ-domain specific. **a**, NHERF2 interacts with the C-terminus of PTH1R. Full-length NHERF2 was overlaid on a membrane containing GST, GST C-terminal tail of PTH1R and mutants lacking 20 C-terminal amino acids (CΔ20), and 20 (NΔ20), 40 (NΔ40), and 60 (NΔ60) N-terminal amino acids (top); Coomassie blue (bottom). **b**, NHERF2 binds to a C-terminal, PTH1R PDZ interaction motif. Full-length NHERF2 was overlaid on a membrane containing PTH1R–GST C-tail (ETVM), and the C-tail with individual alanine substitutions (top, in bold); Coomassie blue (bottom). **c**, PTH1R interacts with PDZ2 of NHERF2. GST C-tail of PTH1R–WT (top) or PTH1R–M591L (bottom) were overlaid on duplicate membranes containing full-length NHERF2, NHERF2–PDZ1, NHERF2–PDZ2 and clone 12A. **d**, NHERF1 binds PTH1R through a PDZ-domain-specific interactions. Full-length NHERF1 was overlaid on a membrane containing PTH1R–GST C-tail (ETVM), and the C-tail with individual alanine substitutions (top); Coomassie blue (bottom).

and a closely related protein, NHERF1, control the major renal Na^+/H^+ exchanger, NHE3 (ref. 14), which is also regulated by PTH¹⁵.

A series of glutathione-S-transferase (GST)–PTH1R C-terminal-tail fusion proteins were synthesized, purified, run on SDS–polyacrylamide gel electrophoresis (PAGE), and overlay with recombinant NHERF2. Full-length PTH1R C-tail proteins and C-tail proteins with truncations of up to 60 N-terminal amino acids interacted strongly with NHERF2, but the PTH1R C-tail with a 20-amino-acid, C-terminal truncation failed to bind (Fig. 1a). When each of the C-terminal PTH1R amino acids, ETVM, was individually replaced with alanine, substitutions at –3 (ATVM), –2 (EAVM), and 0 (ETVA) markedly disrupted binding to NHERF2, but replacement with alanine at –1 (ETAM) had no effect (Fig. 1b).

The GST–PTH1R C-tail fusion protein interacted with full-length NHERF2, PDZ2 of NHERF2, and 12A, but not with the N terminus of NHERF2, which contains PDZ1 (Fig. 1c). Both the β 2-adrenergic and the purinergic receptors preferentially bind to NHERF2–PDZ1 and have a C-terminal leucine¹⁶, so we then asked whether leucine specifies this interaction. Substitution of the C-terminal methionine with leucine did not alter binding to NHERF2, and neither altered binding to PDZ2 nor conferred binding to PDZ1 (Fig. 1c). To our knowledge, a C-terminal methionine is unique among PDZ-domain-binding proteins, although screening a phage peptide library against NHERF–PDZ2 demonstrated that methionine is acceptable at this position¹⁷. Overlay assays of GST–PTH1R C-tail fusion proteins with alanine substitutions of the four C-terminal amino acids showed that full-length PTH1R C-tail also binds NHERF1 by a PDZ-domain-specific interaction (Fig. 1d).

The Chinese hamster fibroblast cell line, PS120, is ideal for assessing functional interactions between PTH1R and NHERF2 because these cells lack expression of detectable PTH1R (data not

shown) and both NHERF1 and NHERF2 (ref. 14), and thus can be studied after transfection with these proteins. Wild-type PTH1R or PTH1Rs lacking a functional PDZ interaction motif, PTH1R-Ala (M591A) and PTH1R–C Δ 20, were expressed in the absence or presence of NHERF2 in PS120 cells, treated with the cross-linker dithiobis(succinimidylpropionate) (DSP), solubilized, precipitated with anti-PTH1R antibodies, and immunoblotted with anti-NHERF2 antibodies. Coprecipitation of NHERF2 with PTH1R was detected only in cells expressing wild-type PTH1R and NHERF2 (Fig. 2a).

PTH signalling was then assessed in PS120 cells expressing various combinations of PTH1R and NHERF2. In cells co-expressing NHERF2 and either PTH1R-Ala or PTH1R–C Δ 20, or in cells expressing wild-type PTH1R in the absence of NHERF2, PTH treatment increased cAMP by 10–15-fold, and inositol phosphates by 2–4 fold (Fig. 2b). In contrast, PTH treatment increased cAMP by only 2-fold in PS120 cells co-expressing the wild-type PTH1R and NHERF2, but increased inositol phosphates by 20–25-fold (Fig. 2b). A minimum of three independently isolated cell lines expressing wild-type PTH1R without or with NHERF2, and PTH1R mutants with NHERF2 were tested with similar results. Specific PTH ligand binding ranged from 20–35%. All cell lines responded equally well to forskolin, and expression levels of the α subunit of G_{s} , wild-type and mutant PTH1Rs and NHERF2 (in cells expressing NHERF2) were comparable (data not shown).

Pertussis-toxin pretreatment of PS120 cells that co-express NHERF2 and PTH1R markedly inhibited PTH activation of PLC (Fig. 2c) and partially restored activation of adenylyl cyclase (data not shown). Thus, PTH stimulates $G_{\text{i/o}}$ proteins when the PTH1R is

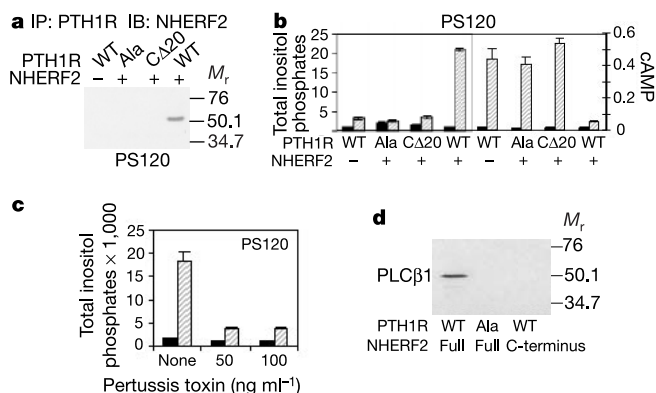


Figure 2 NHERF2 directs PTH1R signalling by assembling PTH1R–PLC β complexes. **a**, NHERF2 co-purifies with PTH1R in mammalian cells. Extracts from PS120 cells expressing PTH1R–WT alone or PTH1R–Ala, PTH1R–C Δ 20 and PTH1R–WT in the presence of NHERF2 were immunoprecipitated with antibodies to PTH1R and detected with NHERF2 antibodies. **b**, NHERF2 mediates PTH1R signalling via PLC. PS120 cells expressing PTH1R–WT alone, or PTH1R–WT, PTH1R–Ala, PTH1R–C Δ 20 and NHERF2 were treated with vehicle (black bars) or 100 nM PTH(1–34) (striped bars) and total inositol phosphates (left) and cAMP (right) determined (mean $\pm$ s.e.m.; $n = 8$). **c**, Pertussis toxin inhibits NHERF2-mediated PTH signalling via PLC. PS120 cells expressing PTH1R–WT and NHERF2 were treated with pertussis toxin (14 h), then with vehicle (black bars) or 100 nM PTH(1–34) (striped bars) and total inositol phosphates determined (mean $\pm$ s.e.m.; $n = 8$). **d**, NHERF2 assembles PTH1R and PLC β 1. GST–PTH1R C-tail WT and GST–PTH1R C-tail-Ala were coupled to glutathione-Sepharose and incubated with full-length NHERF2 or NHERF2 lacking PDZ1 (C terminal). ³⁵S–PLC β 1 C-terminal fragment was then incubated with the Sepharose, the complex eluted with glutathione and analysed by SDS–PAGE and fluorography.

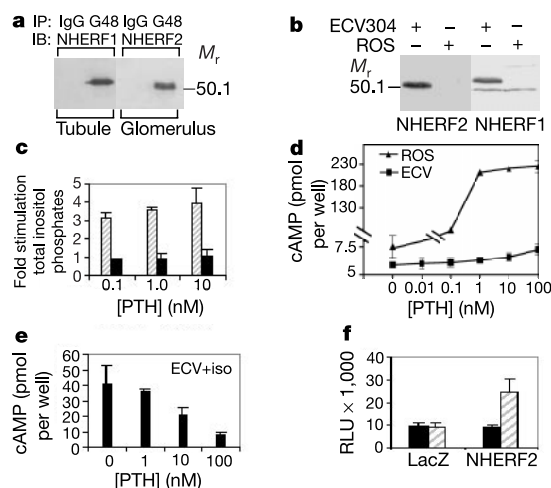


Figure 3 NHERF–PTH1R interactions and signalling in PTH targets. **a**, NHERF1 and NHERF2 co-purify with PTH1R in renal tubular and glomerular extracts. PTH1R immunoprecipitated (IP) with G48 from renal tubular and glomerular extracts were immunoblotted (IB) with NHERF1 and NHERF2 antibodies. **b**, ECV304 cells express NHERF2 and 2, and ROS cells do not. Extracts from ECV304 and ROS cells were immunoblotted with NHERF1 and NHERF2 antibodies. **c**, PTH1R expressed in ECV304 cells, but not ROS cells, activates PLC. ECV304 cells (striped bars) and ROS cells (black bars) were treated with vehicle or PTH(1–34), and total inositol phosphates determined (mean $\pm$ s.e.m.; $n = 8$). **d**, PTH1R expressed in ROS cells, but not ECV304 cells, activates adenylyl cyclase. ECV304 cells (squares) and ROS cells (triangles) were treated with vehicle or PTH(1–34), and cAMP determined (mean $\pm$ s.e.m.; $n = 8$). **e**, Isoproterenol (iso)-stimulated adenylyl cyclase is inhibited by PTH(1–34) in ECV304 cells. ECV304 cells were treated with isoproterenol (50 nM) and vehicle or PTH(1–34), and cAMP determined (mean $\pm$ s.e.m.; $n = 8$). **f**, PTH activation of PKC is mediated by NHERF2. ROS cells, transiently expressing API-luciferase and LacZ or NHERF2 for 24 h, were treated with vehicle (black bars) or PTH(1–34) (striped bars) for 8 h and luciferase activity was analysed (mean $\pm$ s.e.m.; $n = 8$).

bound to NHERF2. Most plausibly, PLC β is activated by $\beta\gamma$ -subunits and adenylyl cyclase is inhibited by α -subunits when they dissociate upon activation of $G_{i/o}$. Incomplete restoration of adenylyl cyclase activation by pertussis toxin suggests that assembly of the PTH1R–NHERF2 complex may sterically hinder coupling to G_s .

All four PLC β isozymes contain consensus C-terminal PDZ interaction motifs. PLC β 1 and β 2 interact with PDZ1 of NHERF1 both *in vitro* and in mammalian cells and tissue¹⁸. We therefore examined whether NHERF2 simultaneously binds PLC β 1 and the PTH1R *in vitro* using GST pull-down assays. A [³⁵S] methionine-labelled, C-terminal fragment of PLC β 1 (relative molecular mass, $M_r \approx 50,000$) and GST–PTH1R C-tail were assembled in the presence of full-length NHERF2 by an interaction that required both an intact PTH1R PDZ-interaction motif and PDZ1 of NHERF2 (Fig. 2d). These data show that NHERF2 forms a multi-protein complex consisting of PLC β 1 bound to PDZ1 and the PTH1R bound to PDZ2.

The potential physiological relevance of PTH signalling through NHERF-assembled PTH1R was investigated in mouse kidney extracts and in PTH-responsive cells that endogenously express PTH1R. Both renal glomeruli and tubules express the PTH1R¹⁹; proximal renal tubules express NHERF1, whereas glomeruli express NHERF2 (ref. 20). As shown in Fig. 3a, NHERF1 co-purifies with PTH1R in extracts enriched with renal tubules, and NHERF2 co-purifies with PTH1R in extracts enriched with glomeruli.

ECV304 cells, a human-endothelium-derived line, and ROS 17/2.8 cells, a rat osteoblastic line, both express functional PTH1R^{21,22} with the same affinity (~ 2 – 5 nM) for PTH(1–34) (data not shown). ECV304 cells, however, abundantly express both NHERF1 and 2, whereas neither NHERF1 nor NHERF2 was detected in extracts of ROS 17/2.8 cells (Fig. 3b). We confirmed the data reported by Isaacs *et al.*²¹ that PTH treatment of ECV304 cells increased inositol phosphates by 3–4-fold (Fig. 3c), but increased cAMP by less than 20% (Fig. 3d). In contrast, PTH increased cAMP in ROS 17/2.8 cells by more than 30-fold (Fig. 3d), but did not increase inositol phosphates (Fig. 3c). Furthermore, PTH(1–34) inhibited isoproterenol-activated adenylyl cyclase in ECV304 cells in a dose-dependent manner, suggesting that PTH1R in these cells is coupled to $G_{i/o}$ (Fig. 3e). Abundant expression of NHERFs in ECV304 cells may account for PTH1R signalling primarily through PLC²¹. ROS 17/2.8 cells transiently transfected with NHERF2 display PTH(1–34)-mediated activation of an activator protein 1 (AP1)-responsive-luciferase reporter gene (Fig. 3f), indicating that NHERF2 is sufficient to couple PTH1R to PKC activation through PLC.

NHERF-mediated signalling provides a new mechanism by which cells regulate responses to PTH and PTHrP. For example, renal proximal tubules express PTH1R on both brush border and basolateral membranes¹⁹, but NHERF1 is expressed only on brush border membrane²⁰. PTH increases cAMP accumulation only on basolateral membrane²³, but it activates protein kinase C and promotes internalization of the sodium-phosphate 2a cotransporter on brush border membrane²⁴. In addition, PTH treatment of opossum kidney (OK) cells, a renal proximal tubule model, activates both PLC β and inhibits phosphate uptake by stimulation of a pertussis-toxin-sensitive G protein⁷. Immunohistochemical analysis suggests that PTH1R¹⁹ and NHERF2²⁰ colocalize in endothelial cells of the peritubular vasculature and NHERF2 is expressed by vascular endothelium in many tissues (data not shown). PTH and PTHrP also cause endothelium-dependent vasodilation²⁵. PTH and/or PTHrP stimulation of NHERF–PTH1R complexes may also explain increases of intracellular calcium and/or inositol phosphates in squamous carcinoma cells²⁶, insulinoma cells²⁷, adult cardiomyocytes²⁸, and lymphocytes²⁹, in which PTH stimulates little or no adenylyl cyclase response. Assembly of G-protein-coupled receptor signalling components by NHERFs or other PDZ-domain proteins,

such as InaD in the *Drosophila eye*³⁰, may be a general mechanism by which cells specify sites of second messenger generation. □

Methods

Complementary DNA constructs

GST fusion proteins of the PTH1R C-terminal tail were generated by polymerase chain reaction (PCR) using *pfu* polymerase (Stratagene). PCR fragments corresponding to full-length C-tail (amino acids 463–591), Δ C20 (463–571), Δ N20 (483–591), Δ N40 (503–591) or Δ N60 (523–591) were ligated into pGEX-5X-1 (Amersham Pharmacia). PTH1R C-tail point mutations were generated by incorporation of the appropriate mutation within the reverse primer. DNA fragments corresponding to full-length NHERF1, and NHERF2, N-terminal/PDZ1 (amino acids 1–100), C-terminal (140–337), PDZ2 (140–240) of NHERF2 and clone 12A were ligated into pT7HisMyc, a bacterial expression plasmid which places a 6XHis and Myc epitope at the N terminus.

Overlay assays

GST fusion proteins were expressed in BL21 *Escherichia coli* and purified using glutathione-Sepharose (Amersham Pharmacia). NHERF1 and 2 proteins were expressed in BL21(DE3)pLysS *E. coli*, and purified using Ni-NTA agarose (Qiagen). Interactors blotted on the membrane were overlaid with the second interactor in a solution containing phosphate buffered saline (PBS) 5% non-fat milk and 0.1% Tween 20 at a concentration of 200 ng ml^{−1}. Interactions were detected with antibodies against the interactor in solution—G48 for PTH1R C-tail and anti-Myc (9e10) for NHERF1 and NHERF2—and visualized using enhanced chemiluminescence (NEN).

Cell culture

PS120 cells were passaged and stably transfected with NHERF2 as described¹⁴. PTH1R-wild-type (WT), PTH1R-Ala (M591A) and PTH1R- Δ C20 were cloned into pIRESpuro2 vector (Clontech) and stably expressed in PS120 cells using FuGene6 (Roche) in the absence or presence of stably expressed NHERF2. Cell lines, selected with 5 μ g ml^{−1} of puromycin, specifically bound 20 to 35% of [¹²⁵I]-PTH(1–34).

Co-immunoprecipitation

PS120 cell lines were treated with 0.25 mM dithiois(succinimidyl)propionate (Pierce) in PBS for 30 min at room temperature, followed by a wash with Tris-buffered saline. Cells were lysed using a solution containing: 25 mM HEPES, pH 7.4, 20% glycerol, 150 mM NaCl, 1 mM EDTA, 0.5% Triton X-100 and a protease inhibitor cocktail (Sigma). Whole-cell extracts were precipitated with PTH1R antibodies (G48) and protein G agarose (Santa Cruz). The pellets were washed with lysis buffer, eluted with SDS sample buffer supplemented with 50 mM dithiothreitol and immunoblotted with NHERF2 antibodies. Isolation of glomerular and tubular extracts from mouse kidney was as described³¹. Briefly, renal cortexes were minced and filtered through a 200- μ m mesh. Tissue that did not pass consisted mainly of tubules and was collected. Glomeruli were trapped by passing the filtrate through a 70- μ m mesh. Glomerular and tubular homogenates were prepared³¹, extracted with lysis buffer and PTH1R immunoprecipitations were performed.

Determination of cAMP and inositol phosphates

Determinations of cAMP and total inositol phosphates were as described¹¹. Values were normalized to PTH1R radioligand binding and total protein per well. Where indicated, cells were incubated with pertussis toxin for 14 h prior to assay.

GST pulldown assay

The C-terminal fragment of rat PLC β 1, corresponding to amino acids 886 to 1216, was cloned into the pT7 expression vector using *pfu* polymerase. PTH1R–GST C-tail and C-tail–M591A were coupled to glutathione-Sepharose and incubated with 200 ng of either full-length NHERF2 or NHERF2–C-terminal fragment for 2 h at 4 °C. The PLC β 1 fragment, *in vitro* translated with TNT reticulocyte lysate (Promega) and [³⁵S]-methionine, was incubated with the Sepharose beads for 4 h at 4 °C in interaction buffer (24 mM HEPES, pH 7.4, 20% glycerol, 150 mM NaCl). The beads were then washed with interaction buffer, eluted with 5 mM glutathione and the eluates were analysed by fluorography.

Luciferase reporter gene assay

An AP1-luciferase reporter gene (Clontech) (50 ng per well) and either pcDNA3.1–LacZ or pcDNA3.1–NHERF2 (1 μ g per well) were transiently expressed in ROS 17/2.8 cells using FuGene 6 in serum-free media. Twenty-four hours post-transfection, cells were treated with vehicle or 100 nM PTH(1–34) for 8 h and luciferase activity was analysed using the Bright-Glo luciferase assay system (Promega). Transfection efficiency was assessed with a human growth hormone expression plasmid and media analysed by RIA (Nichols Diagnostics).

Received 2 November 2001; accepted 8 April 2002; doi:10.1038/nature00816.

1. Nissenson, R. In *The Parathyroids: Basic and Clinical Concepts*, 2nd edn (eds Bilezikian, J., Marcus, R. & Levine, M.) 93–104 (Academic, San Diego, 2001).
2. Strewler, G. J. In *The Parathyroids: Basic and Clinical Concepts*, 2nd edn (eds Bilezikian, J., Marcus, R. & Levine, M.) 213–226 (Academic, San Diego, 2001).
3. Bringham, F. R. In *The Parathyroids: Basic and Clinical Concepts*, 2nd edn (eds Bilezikian, J., Marcus, R. & Levine, M.) 227–244 (Academic, San Diego, 2001).
4. Segre, G. & Lee, K. In *The Parathyroids: Basic and Clinical Concepts*, 2nd edn (eds Bilezikian, J., Marcus, R. & Levine, M.) 245–260 (Academic, San Diego, 2001).
5. Clemons, T. L. & Broadus, A. E. In *The Parathyroids: Basic and Clinical Concepts*, 2nd edn (eds

- Bilezikian, J., Marcus, R. & Levine, M. 261–274 (Academic, San Diego, 2001).
6. Abou-Samra, A. B. *et al.* Expression cloning of a common receptor for parathyroid hormone and parathyroid hormone-related peptide from rat osteoblast-like cells: A single receptor stimulates intracellular accumulation of both cAMP and inositol triphosphates and increases intracellular free calcium. *Proc. Natl Acad. Sci. USA* **89**, 2732–2736 (1992).
 7. Reshkin, S. J., Forgo, J. & Murer, H. Apical and basolateral effects of PTH in OK cells: transport inhibition, messenger production, effects of pertussis toxin, and interaction with a PTH analog. *J. Membr. Biol.* **124**, 227–237 (1991).
 8. Tanaka, H., Smogorzewski, M., Koss, M. & Massry, S. G. Pathways involved in PTH-induced rise in cytosolic Ca^{2+} concentration of rat renal proximal tubule. *Am. J. Physiol.* **268**, F330–F337 (1995).
 9. Smogorzewski, M., Zayed, M., Zhang, Y. B., Roe, J. & Massry, S. G. Parathyroid hormone increases cytosolic calcium concentration in adult rat cardiac myocytes. *Am. J. Physiol.* **264**, H1998–H2006 (1993).
 10. Miyachi, A. *et al.* Stimulation of transient elevations in cytosolic Ca^{2+} is related to inhibition of PI transport in OK cells. *Am. J. Physiol.* **259**, F485–F493 (1990).
 11. Iida-Klein, A. *et al.* Truncation of the carboxyl-terminal region of the rat parathyroid hormone (PTH)/PTH-related peptide receptor enhances PTH stimulation of adenyl cyclase but not phospholipase C. *J. Biol. Chem.* **270**, 8458–8465 (1995).
 12. Doyle, D. A. *et al.* Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. *Cell* **85**, 1067–1076 (1996).
 13. Bretscher, A., Chambers, D., Nguyen, R. & Rezek, D. ERM-Merlin and EBP50 protein families in plasma membrane organization and function. *Annu. Rev. Cell Dev. Biol.* **16**, 113–143 (2000).
 14. Yun, C. H. *et al.* cAMP-mediated inhibition of the epithelial brush border Na^{+}/H^{+} exchanger, NHE3, requires an associated regulatory protein. *Proc. Natl Acad. Sci. USA* **94**, 3010–3015 (1997).
 15. Krayer-Pawlowska, D., Helmle-Kolb, C., Montrose, M. H., Krapf, R. & Murer, H. Studies on the kinetics of Na^{+}/H^{+} exchange in OK cells: introduction of a new device for the analysis of polarized transport in cultured epithelia. *J. Membr. Biol.* **120**, 173–183 (1991).
 16. Hall, R. A. *et al.* A C-terminal motif found in the beta2-adrenergic receptor, P2Y1 receptor and cystic fibrosis transmembrane conductance regulator determines binding to the Na^{+}/H^{+} exchanger regulatory factor family of PDZ proteins. *Proc. Natl Acad. Sci. USA* **95**, 8496–8501 (1998).
 17. Wang, S., Raab, R. W., Schatz, P. J., Guggino, W. B. & Li, M. Peptide binding consensus of the NHERF-PDZ1 domain matches the C-terminal sequence of cystic fibrosis transmembrane conductance regulator (CFTR). *FEBS Lett.* **427**, 103–108 (1998).
 18. Tang, Y. *et al.* Association of mammalian trp4 and phospholipase C isozymes with a PDZ domain-containing protein, NHERF. *J. Biol. Chem.* **275**, 37559–37564 (2000).
 19. Amizuka, N. *et al.* Cell-specific expression of the parathyroid hormone (PTH)/PTH-related peptide receptor gene in kidney from kidney-specific and ubiquitous promoters. *Endocrinology* **138**, 469–481 (1997).
 20. Wade, J. B., Welling, P. A., Donowitz, M., Shenolikar, S. & Weinman, E. J. Differential renal distribution of NHERF isoforms and their colocalization with NHE3, ezrin, and ROMK. *Am. J. Physiol. Cell Physiol.* **280**, C192–C198 (2001).
 21. Isales, C. M. *et al.* Functional parathyroid hormone receptors are present in an umbilical vein endothelial cell line. *Am. J. Physiol. Endocrinol. Metab.* **279**, E654–E662 (2000).
 22. Yamamoto, I., Shigeno, C., Potts, J. T. Jr & Segre, G. V. Characterization and agonist-induced down-regulation of parathyroid hormone receptors in clonal rat osteosarcoma cells. *Endocrinology* **122**, 1208–1217 (1988).
 23. Kaufmann, M. *et al.* Apical and basolateral parathyroid hormone receptors in rat renal cortical membranes. *Endocrinology* **134**, 1173–1178 (1994).
 24. Traebert, M., Volkl, H., Biber, J., Murer, H. & Kaissling, B. Luminal and contraluminal action of 1–34 and 3–34 PTH peptides on renal type IIa Na-P(i) cotransporter. *Am. J. Physiol. Renal Physiol.* **278**, F792–F798 (2000).
 25. Sutliff, R. L. *et al.* Vasorelaxant properties of parathyroid hormone-related protein in the mouse: evidence for endothelium involvement independent of nitric oxide formation. *Endocrinology* **140**, 2077–2083 (1999).
 26. Orloff, J. J. *et al.* Analysis of PTHRP binding and signal transduction mechanisms in benign and malignant squamous cells. *Am. J. Physiol.* **262**, E599–E607 (1992).
 27. Gaich, G. *et al.* Amino-terminal parathyroid hormone-related protein: specific binding and cytosolic calcium responses in rat insulinoma cells. *Endocrinology* **132**, 1402–1409 (1993).
 28. Schluter, K. D., Weber, M. & Piper, H. M. Parathyroid hormone induces protein kinase C but not adenylate cyclase in adult cardiomyocytes and regulates cyclic AMP levels via protein kinase C-dependent phosphodiesterase activity. *Biochem. J.* **310**, 439–444 (1995).
 29. McCauley, L. K., Rosol, T. J., Merryman, J. I. & Capen, C. C. Parathyroid hormone-related protein binding to human T-cell lymphotropic virus type I-infected lymphocytes. *Endocrinology* **130**, 300–306 (1992).
 30. Tsunoda, S. *et al.* A multivalent PDZ-domain protein assembles signalling complexes in a G-protein-coupled cascade. *Nature* **388**, 243–249 (1997).
 31. Kerjaschki, D. & Farquhar, M. G. Immunocytochemical localization of the Heymann nephritis antigen (GP330) in glomerular epithelial cells of normal Lewis rats. *J. Exp. Med.* **157**, 667–686 (1983).

Acknowledgements

We thank D. Moore for providing the pT7-hismyc vector; S. Goo Rhee for providing the cDNA for PLC β 1; V. Ramesh for the NHERF1 cDNA; and C. Isales for providing the ECV304 cells. We also would like to thank J. Potts and H. Kronenberg for their helpful discussions and review of the manuscript. This work was supported in part by the NIH (G.V.S.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.V.S. (e-mail: segre@helix.mgh.harvard.edu).

TNF-mediated inflammatory skin disease in mice with epidermis-specific deletion of IKK2

Manolis Pasparakis^{*†}, Gilles Courtois[‡], Martin Hafner^{§||}, Marc Schmidt-Suppran^{*||}, Arianna Nenci[†], Atiye Toksoy^{||}, Monika Krampert[#], Matthias Goebeler[¶], Reinhard Gillitzer[¶], Alain Israel[‡], Thomas Krieg[§], Klaus Rajewsky^{*||} & Ingo Haase[§]

^{*} Institute for Genetics, University of Cologne, Weyertal 121, D-50931 Cologne, Germany

[†] EMBL Mouse Biology Programme, via Ramarini 32, I-00016 Monterotondo, Italy

[‡] Unité de Biologie Moléculaire de l'Expression Génique, URA 1773 CNRS, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France

[§] Department of Dermatology and Center for Molecular Medicine, University of Cologne, Joseph-Stelzmann Strasse 9, D-50924 Cologne, Germany

[¶] Department of Dermatology, University of Würzburg, Josef-Schneider Strasse 2, 97080 Würzburg, Germany

[#] Institute for Cell Biology, ETH Höggerberg, HPM D25, CH-8093 Zürich, Switzerland

The I κ B kinase (IKK), consisting of the IKK1 and IKK2 catalytic subunits and the NEMO (also known as IKK γ) regulatory subunit, phosphorylates I κ B proteins, targeting them for degradation and thus inducing activation of NF- κ B (reviewed in refs 1, 2). IKK2 and NEMO are necessary for NF- κ B activation through pro-inflammatory signals^{3–8}. IKK1 seems to be dispensable for this function but controls epidermal differentiation independently of NF- κ B^{9–12}. Previous studies suggested that NF- κ B has a function in the growth regulation of epidermal keratinocytes^{12–14}. Mice lacking RelB or I κ B α , as well as both mice and humans with heterozygous NEMO mutations, develop skin lesions^{7,8,15–18}. However, the function of NF- κ B in the epidermis remains unclear¹⁹. Here we used Cre/loxP-mediated gene targeting to investigate the function of IKK2 specifically in epidermal keratinocytes. IKK2 deficiency inhibits NF- κ B activation, but does not lead to cell-autonomous hyperproliferation or impaired differentiation of keratinocytes. Mice with epidermis-specific deletion of IKK2 develop a severe inflammatory skin disease, which is caused by a tumour necrosis factor-mediated, $\alpha\beta$ T-cell-independent inflammatory response that develops in the skin shortly after birth. Our results suggest that the critical function of IKK2-mediated NF- κ B activity in epidermal keratinocytes is to regulate mechanisms that maintain the immune homeostasis of the skin.

Mice lacking IKK2 die *in utero* as a result of tumour necrosis factor (TNF)-induced hepatocyte apoptosis^{3–5}. To study the function of IKK2 in different tissues we generated mice carrying a loxP-site-flanked (floxed (FL)) *Ikk2* allele (Fig. 1a). *Ikk2*^{FL/FL} mice develop normally and express normal levels of IKK2 (Fig. 1b). Excision of the floxed *Ikk2* sequences by crossing to a Cre-deleter transgenic strain²⁰ produced mice carrying deleted (D) *Ikk2* alleles. As expected, *Ikk2*^{D/D} mice die *in utero* and exhibit liver degeneration (data not shown). Analysis of mouse embryonic fibroblasts (MEFs) revealed that *Ikk2*^{D/D} cells lack expression of IKK2 (Fig. 1b) and show impaired activation of NF- κ B on stimulation with TNF, interleukin (IL)-1 or bacterial lipopolysaccharide (LPS) (ref. 7 and data not shown). To ablate IKK2 specifically in the epidermis we used a transgenic mouse line expressing Cre under the control of the human keratin 14 (K14) promoter. LacZ staining of tissue

|| Present addresses: Center for Blood Research, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA (M.S.S. and K.R.); German Research Centre for Biotechnology, Mascheroder Weg 1, 38124 Braunschweig, Germany (M.H.).

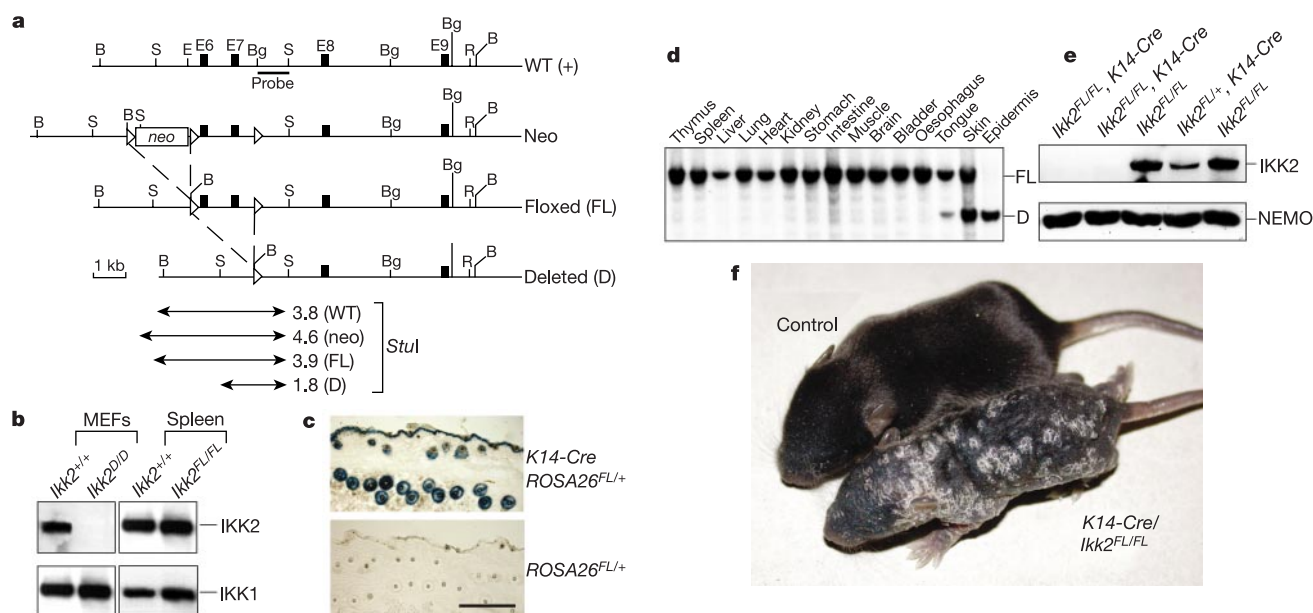


Figure 1 Epidermis-specific knockout of IKK2. **a**, Diagram showing the wild-type *Ikk2* genomic locus (WT), the neomycin-resistance-containing (neo), the loxP-flanked (FL) and the deleted (D) *Ikk2* alleles. Filled boxes indicate exons (E6–E9). Restriction enzyme sites and the location of the probe used for Southern blot analysis are depicted. *StuI* fragments are in kilobases. B, *Bam*HI; E, *Eco*RI; Bg, *Bgl*II; S, *Stu*I. LoxP sites are indicated by arrowheads. **b**, Western blot analysis of IKK2 and IKK1 expression in wild-type and *Ikk2*^{FL/FL} MEFs and in splenocytes from wild-type and *Ikk2*^{FL/FL} mice. **c**, Analysis of β -galactosidase

expression in the skin of a *K14-Cre, ROSA26*^{FL/+} mouse shows Cre recombination specifically in the epidermis and in hair follicles. Scale bar, 200 μ m. **d**, Tissue specificity of IKK2 inactivation. DNA isolated from various tissues of a *K14-Cre/Ikk2*^{FL/FL} mouse was subjected to Southern blot analysis after digestion with *StuI*. Deletion is detected in DNA from epidermis, total skin and tongue. **e**, Western blot analysis of IKK2 and NEMO expression in the epidermis of newborn mice with the indicated genotypes. **f**, *K14-Cre/Ikk2*^{FL/FL} and control mice at P8.

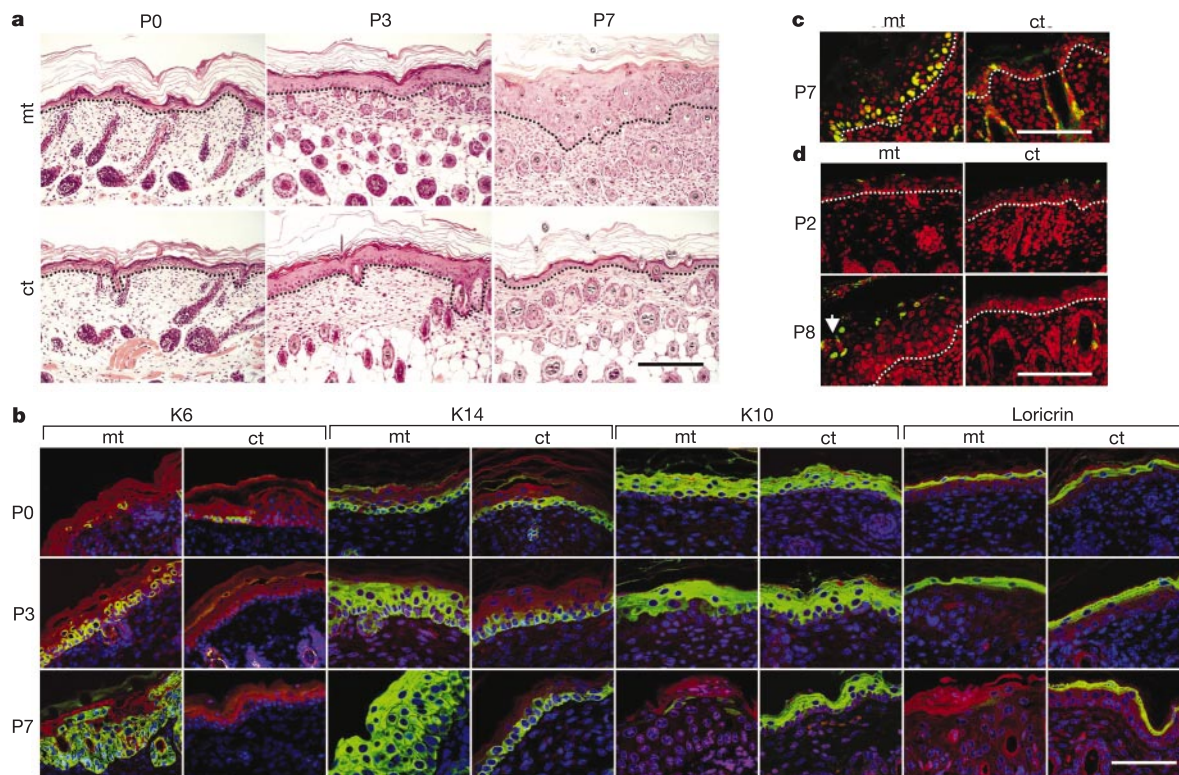


Figure 2 Histology of the skin in mice with epidermis-specific IKK2 deficiency. **a**, **b**, Skin sections from *K14-Cre/Ikk2*^{FL/FL} (mutant, mt) and control (ct) mice were stained with haematoxylin/eosin (**a**) or immunostained for the indicated epidermal differentiation markers (green staining in **b**). Red counterstaining in **b** shows tissue structure; blue staining shows nuclei. **c**, Analysis of cell proliferation by staining for Ki67 (green); red

staining shows nuclei. **d**, TUNEL staining for the detection of apoptotic cells (green); red staining shows nuclei. A white arrow points to an intraepidermal pustule with surrounding apoptotic keratinocytes. Dotted lines indicate the border between epidermis and dermis. Scale bars: **a**, 150 μ m; **b**, 100 μ m; **c**, **d**, 200 μ m.

sections from a *K14-Cre* mouse carrying a *ROSA26^{FL}* Cre-reporter allele²¹ (Fig. 1c and data not shown) and Southern blot analysis of DNA isolated from various tissues of a *K14-Cre/Ikk2^{FL/FL}* mouse (Fig. 1d) showed Cre recombination specifically in the epidermis, hair follicles and the epithelium of the tongue. Western analysis of epidermal extracts prepared from newborn pups failed to detect expression of IKK2 in *K14-Cre/Ikk2^{FL/FL}* mice, demonstrating that IKK2 protein is absent from the epidermis at birth (Fig. 1e).

K14-Cre/Ikk2^{FL/FL} pups are born at the expected mendelian ratios and are macroscopically indistinguishable from their littermates until postnatal day 4–5 (P4–P5), when their skin starts becoming hard and inflexible. This phenotype progresses rapidly and by P7–P8 the mice exhibit a highly rigid, shell-like skin with widespread scaling (Fig. 1f). At this stage the mice become ‘runted’ and they subsequently die between P7 and P9. The cause of death is not known at present; however, it does not seem to be the result of impaired skin barrier function as no differences were detected in toluidine blue penetration into the skin between mutant and control mice at P8 (data not shown). *Ikk2^{FL/FL}* mice lacking the *K14-Cre* transgene and *Ikk2^{FL/+}* and *Ikk2^{+/+}* mice with or without *K14-Cre* did not show any pathological skin phenotype and are collectively referred to as controls. Histological analysis of skin sections from *K14-Cre/Ikk2^{FL/FL}* mice at P7 revealed a markedly thickened epidermis with loss of the granular layer, pronounced hyperkeratosis, focal parakeratosis, subcorneal pustule formation and increased cellularity and dilated blood vessels in the dermis (Fig. 2a).

We undertook detailed immunohistological analysis of skin sections from *K14-Cre/Ikk2^{FL/FL}* and control mice at different time points after birth, using antibodies against various epidermal

differentiation markers. At P0 and P1 the skin of mutant mice did not show any differences compared to controls (Fig. 2b, and data not shown), suggesting that, in contrast to IKK1, IKK2 is not essential for epidermal differentiation. At P3 the epidermis of *K14-Cre/Ikk2^{FL/FL}* mice showed upregulation of keratin 6 (K6) expression, a marker for inflamed and hyperproliferative epidermis, and suprabasal expression of K14, which is normally confined to the basal epidermal layer (Fig. 2b). K6 and K14 were expressed in all viable layers of the mutant epidermis at P7 (Fig. 2b). In addition, expression of the suprabasal keratin 10 (K10) and of the keratinocyte terminal differentiation markers loricrin (Fig. 2b) and filaggrin (not shown) was markedly reduced in *K14-Cre/Ikk2^{FL/FL}* animals at this stage. Staining for Ki67 revealed increased proliferation of keratinocytes in mutant epidermis at P4 and P7, but not P0 and P1 (Fig. 2c, and data not shown). TdT-mediated dUTP nick end labelling (TUNEL) staining showed no differences between *K14-Cre/Ikk2^{FL/FL}* and control skin at P2 (Fig. 2d); however, increased numbers of apoptotic keratinocytes in proximity to aggregates of inflammatory cells were observed in mutant epidermis at P8 (Fig. 2d).

To determine whether the cutaneous phenotype observed in *K14-Cre/Ikk2^{FL/FL}* mice is caused by a cell-autonomous defect in proliferation and/or differentiation of IKK2-deficient keratinocytes, we analysed *ex vivo* keratinocyte cultures from mutant and control mice. Keratinocytes isolated from newborn *K14-Cre/Ikk2^{FL/FL}* mice lack expression of IKK2 (Fig. 3a) and, similarly to *Ikk2*-knockout MEFs^{3–5,7}, show impaired NF- κ B activation in response to stimulation by TNF or IL-1 β (Fig. 3b, c). IKK2-deficient keratinocytes were not impaired in their capability to terminally differentiate in suspension culture, as demonstrated by similar expression of K10, involucrin, loricrin and filaggrin compared to control keratinocytes

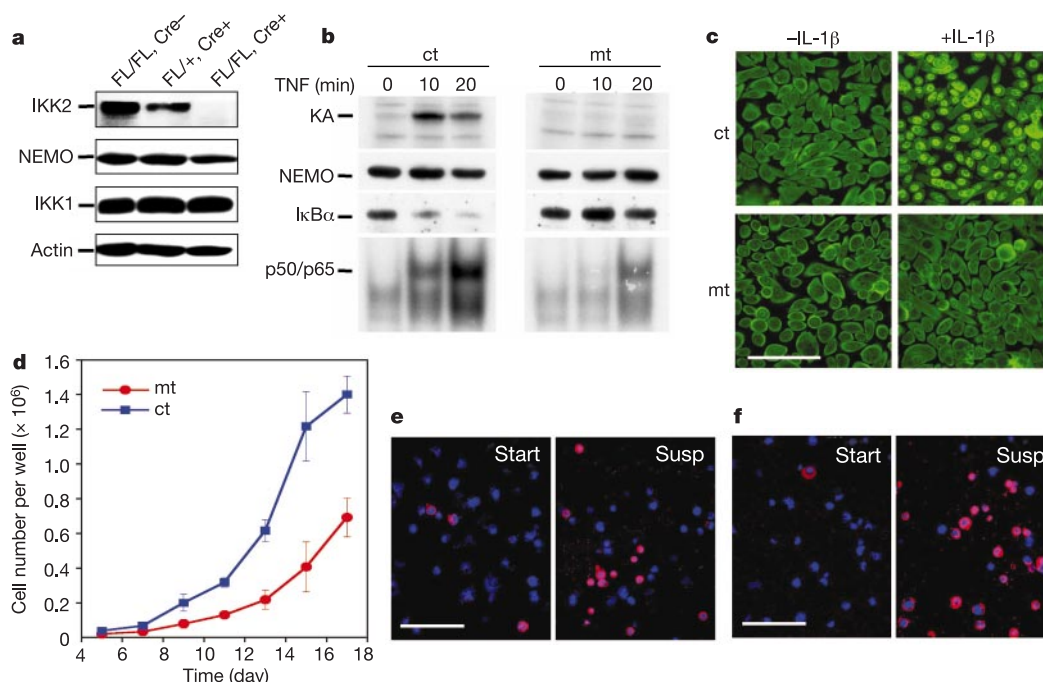


Figure 3 Analysis of NF- κ B activation, proliferation and differentiation in cultured keratinocytes. **a**, Western blot analysis showing expression of IKK2, IKK1, NEMO and actin in keratinocytes isolated from mice with the indicated genotypes. **b**, I κ B kinase activity (KA), I κ B α degradation and NF- κ B activation in primary keratinocytes isolated from control (ct) and *K14-Cre/Ikk2^{FL/FL}* (mt) mice. Cells were treated with TNF (20 ng ml⁻¹) for the indicated times. The kinase activity of IKK complexes immunoprecipitated with anti-NEMO antibody was measured. NEMO blot is shown as loading control. I κ B α degradation was assayed by western analysis of cytoplasmic extracts and NF- κ B DNA-binding activity was measured by gel mobility shift analysis of nuclear extracts. **c**, Immunostaining for p65/

RelA in unstimulated or IL-1 β -treated (20 ng ml⁻¹ for 20 min) mutant and control keratinocytes. **d**, Growth curves of primary keratinocytes in culture. A total of 30,000 cells isolated from mutant (red circles) and control mice (blue squares) were seeded per well, cultured for 17 days and collected at different time points. The graph shows mean values $\pm$ s.d. from triplicate wells. **e**, **f**, Expression of differentiation markers in IKK2-deficient primary keratinocytes before (start) or after 24 h in suspension culture (susp). Fluorescence images of keratinocytes expressing keratin 10 (red in **e**) and involucrin (red in **f**) are shown. Blue signal shows nuclei stained with 4,6-diamidino-2-phenylindole (DAPI). Scale bars: **c**, **e**, **f**, 200 μ m.

(Fig. 3e–f, see also Supplementary Information Table 1). Growth curves revealed that primary keratinocytes lacking IKK2 exhibit decreased proliferation compared with control cells (Fig. 3d). These findings show that, in contrast to IKK1, IKK2 deficiency does not lead to impaired differentiation and hyperproliferation of keratinocytes. Previous *in vivo* and *in vitro* experiments showed hyperproliferation and epidermal hyperplasia after NF- κ B inhibition and growth arrest after NF- κ B activation in epidermal keratinocytes, and suggested a role for NF- κ B in the regulation of keratinocyte growth^{12–14}. In these studies NF- κ B activity in keratinocytes was

modulated downstream of the IKK complex. Our experiments demonstrate that the proposed growth regulatory function of NF- κ B in epidermal keratinocytes can be mediated through an IKK2-independent pathway and suggest that the skin phenotype of *K14-Cre/Ikk2^{FL/FL}* mice is not caused by cell-autonomous hyperproliferation of IKK2-deficient keratinocytes. TUNEL staining revealed that keratinocytes lacking IKK2 do not undergo increased spontaneous apoptosis in culture and show only mildly increased apoptosis on TNF treatment (1.9–2.8% compared with 0.1–0.2% in the controls). In agreement with the presence of only a few

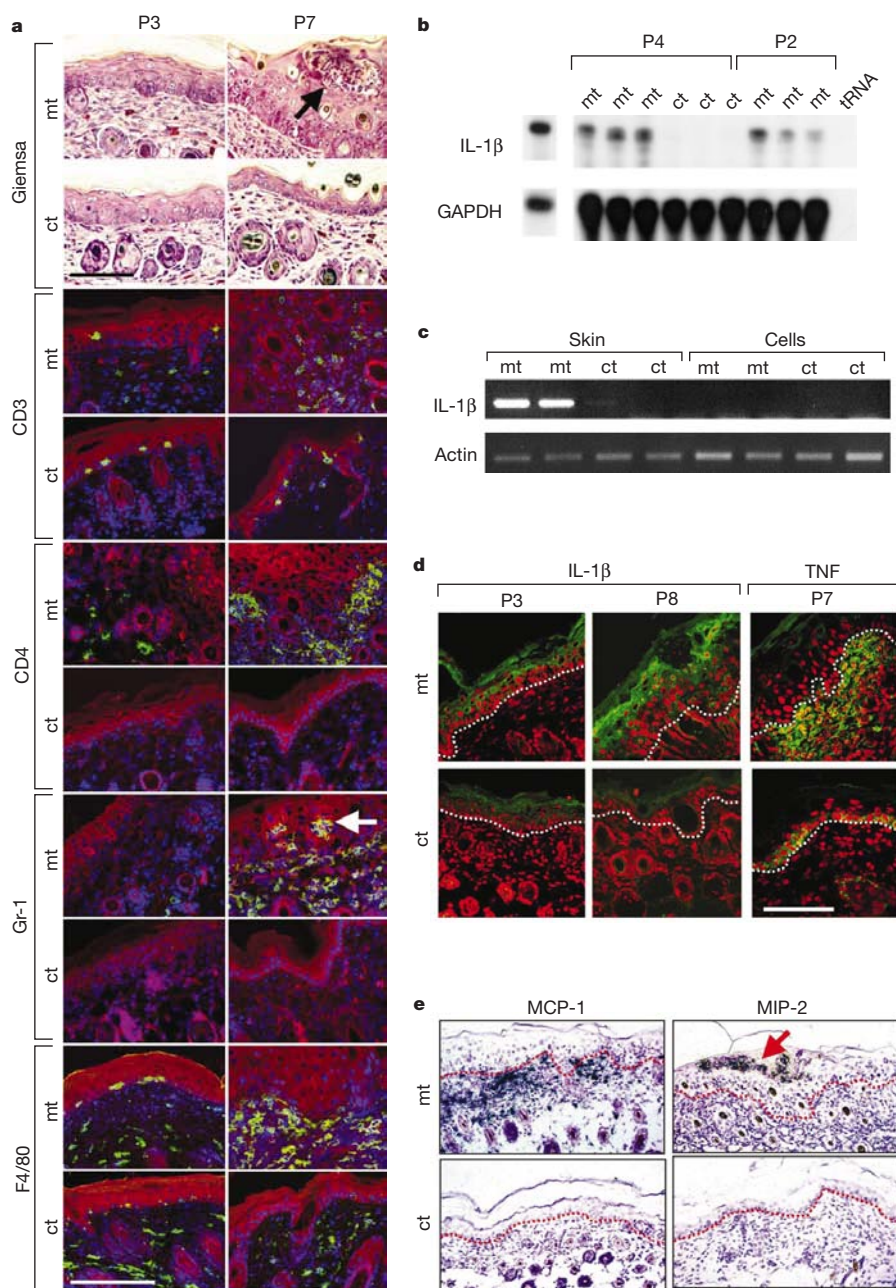


Figure 4 Inflammation in the skin of *K14-Cre/Ikk2^{FL/FL}* mice. **a**, Sections were stained with Giemsa or immunostained with antibodies recognizing T cells (CD3, CD4), granulocytes (Gr-1) and macrophages (F4/80) (green signal). Counterstaining and nuclear staining is as in Fig. 2b. Arrows point to intraepidermal pustules. **b**, RNase protection assay showing IL-1 β expression in the skin of mice 2 or 4 days after birth. Left lane shows positive controls; tRNA is used as negative control. **c**, RT-PCR analysis of mouse IL-1 β expression in skin samples from mutant and control mice at P4 and in cultured IKK2-

deficient (mt) and control (ct) keratinocytes (cells). **d**, Immunostaining using antibodies specific for either IL-1 β or TNF (green) in skin sections from mice at P3, and P7–P8; red staining shows nuclei. **e**, *In situ* hybridization for MCP1 and MIP-2 mRNA (black) in skin sections from mice at P4 (MCP-1) or P7 (MIP-2). mt, mutant; ct, control mice. Magnification $\times 250$. Dotted lines indicate the border between epidermis and dermis. Scale bars, **a**, **d**, 200 μ m.

apoptotic keratinocytes in the epidermis of *K14-Cre/Ikk2^{FL/FL}* mice, these results suggest that TNF-induced killing of IKK2 knockout keratinocytes does not have an important role in the pathogenesis of the skin phenotype in our mice.

The progressive disturbance of epidermal differentiation and proliferation in the skin of *K14-Cre/Ikk2^{FL/FL}* mice seems to be associated with an inflammatory response, suggested by the presence of subcorneal pustules (Fig. 4a) and increased cellularity in the dermis at P7. Immunohistological analysis revealed the presence of increased numbers of macrophages, granulocytes and CD4-positive T cells in the dermis of mutant mice at P7 (Fig. 4a). Examination of inflammatory cytokine gene expression showed increased levels of IL-1 β messenger RNA in the skin of *K14-Cre/Ikk2^{FL/FL}* mice as early as P2, but not at P1 (Fig. 4b and data not shown). Curiously, IKK2-deficient epidermal keratinocytes are identified as the major source of these elevated IL-1 β levels, as demonstrated by immunostaining with an antibody against IL-1 β (Fig. 4d) and by analysis of IL-1 β

mRNA in epidermal sheets (data not shown). However, the increased expression of IL-1 β does not seem to be induced by a cell-autonomous mechanism triggered by the lack of IKK2, as cultured IKK2-deficient keratinocytes do not show increased expression of IL-1 β (Fig. 4c and data not shown). Immunohistology with anti-TNF antibodies revealed increased expression of TNF by cells in the dermis—most probably macrophages and granulocytes—of *K14-Cre/Ikk2^{FL/FL}* mice at P4 and P7, but not at P1 and P2 (Fig. 4d and data not shown). To identify the signals that attract the inflammatory cells into the skin we performed *in situ* hybridization analysis for various chemokine mRNAs. These experiments showed expression of MCP-1 (Fig. 4e), LIX, C10, IP-10 and Mig (data not shown) in the dermis of *K14-Cre/Ikk2^{FL/FL}* mice at P4 and P7, but not in the controls. MIP-2, one of the murine homologues of human IL-8, was expressed mainly in pustules of inflammatory cells in the epidermis of *K14-Cre/Ikk2^{FL/FL}* mice at P7 (Fig. 4e). These results suggest that an immune response orchestrated by the expression of inflammatory cytokines and multiple chemokines in the dermis, and involving T lymphocytes, macrophages and granulocytes, is implicated in the pathogenesis of the skin disease that develops in these mice.

To investigate whether $\alpha\beta$ T lymphocytes are essential for the development of the skin disease we crossed the *K14-Cre/Ikk2^{FL/FL}* mice with T-cell antigen receptor- α (TCR α) knockout mice, which lack mature $\alpha\beta$ T cells²². *K14-Cre/Ikk2^{FL/FL}/TCR α ^{-/-}* mice develop skin disease with similar kinetics and histological characteristics as the *K14-Cre/Ikk2^{FL/FL}* mice (Fig. 5b), demonstrating that $\alpha\beta$ T cells are not required for the development of the inflammatory skin lesions. This finding, together with the early age of disease onset, suggests that the inflammation of the skin in *K14-Cre/Ikk2^{FL/FL}* mice is not triggered by an antigen-specific immune response but rather by a reaction driven from the innate immune system.

To investigate the role of TNF signalling in the pathogenesis of the skin disease we crossed the *K14-Cre/Ikk2^{FL/FL}* mice with TNF-receptor I (TNFRI)-deficient mice²³. Surprisingly, *K14-Cre/Ikk2^{FL/FL}* mice lacking TNFRI do not develop inflammatory skin disease (Fig. 5a). Immunohistological analysis of skin sections from *K14-Cre/Ikk2^{FL/FL}/Tnfr1^{-/-}* mice revealed a normal pattern of keratinocyte differentiation and proliferation in the epidermis and the absence of inflammatory cells from the dermis (Fig. 5b). These results demonstrate that TNFRI signalling is critical for the inflammatory response that causes the skin disease in *K14-Cre/Ikk2^{FL/FL}* mice. As TNFRI mediates its effects largely through activation of NF- κ B-dependent gene transcription, the presence of an intact NF- κ B signalling pathway in all other cell types except epidermal keratinocytes in *K14-Cre/Ikk2^{FL/FL}* mice may be critical for disease pathogenesis. Finally, the normal development of the epidermis in *K14-Cre/Ikk2^{FL/FL}/Tnfr1^{-/-}* mice confirms that IKK2 deficiency does not interfere directly with keratinocyte differentiation and proliferation, and suggests that the observed hyperplasia and the disturbed expression of keratinocyte differentiation markers in the mutant epidermis after P3 are secondary to the inflammation of the skin.

How can deletion of IKK2 in epidermal keratinocytes initiate an inflammatory skin disease? A tightly regulated microenvironment of cytokines and growth factors produced by keratinocytes, dermal fibroblasts and immune cells is critical for the maintenance of physiological skin homeostasis²⁴. The NF- κ B signalling pathway may be essential for epidermal keratinocytes to respond to such regulatory mediators and/or to produce factors that are important for the maintenance of a well balanced interplay between the epidermis, the dermis and the immune system. Although the exact mechanism of disease initiation remains unclear, ablation of IKK2 from epidermal keratinocytes seems to interfere with this balance and triggers an inflammatory response that is marked by the induction of inflammatory cytokines such as IL-1 β and TNF, resulting in the development of skin disease in *K14-Cre/Ikk2^{FL/FL}*

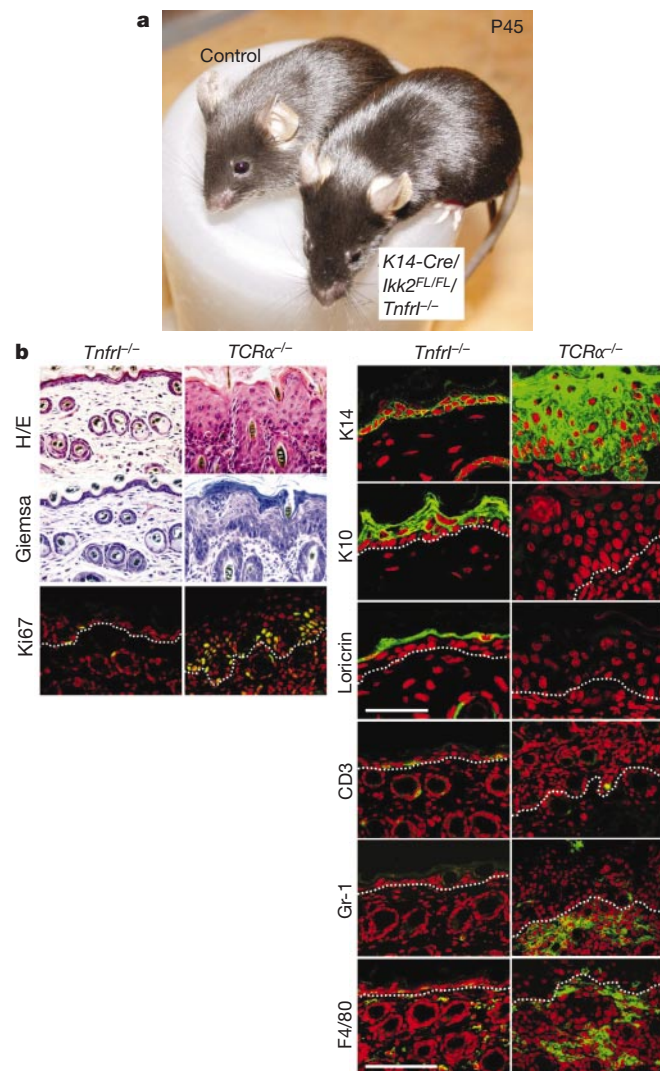


Figure 5 Requirement of TNFRI for the development of the skin phenotype in *K14-Cre/Ikk2^{FL/FL}* mice. **a**, *K14-Cre/Ikk2^{FL/FL}/Tnfr1^{-/-}* mice do not develop inflammatory skin disease. **b**, Analysis of skin samples from *K14-Cre/Ikk2^{FL/FL}/Tnfr1^{-/-}* and *K14-Cre/Ikk2^{FL/FL}/TCR α ^{-/-}* mice. Sections were stained with haematoxylin/eosin (H/E) or Giemsa, or with antibodies against the indicated differentiation, proliferation or immune cell markers (green signal; see Fig. 4 for abbreviations). Nuclei are in red. Top scale bar, 100 μ m; bottom scale bar, 200 μ m. Dotted lines indicate the border between epidermis and dermis.

mice. The neutralization of the skin disease in *K14-Cre/Ikk2^{FL/FL}* mice lacking TNFRI demonstrates the essential pathogenic role of TNF for the development of this phenotype. This result does not exclude a function for IL-1 in our model, but it confirms the role of TNF as a master cytokine in inflammation and highlights its significance in the pathogenesis of inflammatory skin disease. The initiation of the skin disease during the first few days after birth may indicate the involvement of environmental factors. Alternatively, mechanisms related to developmental changes of the skin at this early age might be involved in disease initiation. Our finding that interference with NF- κ B signalling in epidermal keratinocytes triggers skin inflammation provides new insight into the function of NF- κ B in the epidermis and suggests that keratinocytes can act as the initiating cell type in inflammatory skin disease. □

Methods

Gene targeting and transgenic mice

A clone containing the mouse *Ikk2* genomic locus was isolated by screening a P1 embryonic stem cell mouse library (Genome Systems). A 12-kilobase (kb) *Bam*HI fragment containing exons 6–9 of the *Ikk2* gene was mapped and sequenced. The IKK2 targeting vector was constructed using the pEasyFlox plasmid (provided by M. Alimzhanov) by placing a 2-kb *Eco*RI–*Bgl*II genomic fragment, containing exons 6 and 7, between the loxP-flanked PGKneo cassette and the third loxP site. An upstream 2.6-kb *Scal*–*Eco*RI fragment and a 4.2-kb downstream *Bgl*II fragment were used as arms for homology. Bruce-4 embryonic stem cells derived from C57BL/6 mice²⁵ were cultured, transfected and selected as described previously⁷. Homologous recombinant clones were isolated and the loxP-flanked PGKneo cassette was excised by transient expression of Cre recombinase. Germ-line transmitting chimaeras were generated using embryonic stem cells carrying the floxed *Ikk2* gene. Deletion of exons 6 and 7, which correspond to nucleotides 389–567 of the coding IKK2 complementary DNA sequence, introduces premature termination codons producing an *Ikk2* null allele. The *K14-Cre* transgenic mouse will be described elsewhere (M.H., manuscript in preparation). TNFRI-deficient mice were provided by K. Pfeffer.

Western blotting, IKK and NF- κ B assays

Immunoblot analysis, IKK kinase assays and NF- κ B gel mobility shift assays of cytoplasmic and nuclear extracts from cultured keratinocytes were done as described^{7,26}.

Histology and immunostaining

Immunostainings were carried out on paraffin or cryostat sections using polyclonal antibodies against mouse loricrin, filaggrin, K14 and K10 (BabCo) and involucrin, or monoclonal antibodies against mouse CD3, CD8 (Chemicon), CD4 (clone GK1.5/4), Gr-1 (Ly-6G, clone RB6-8C5), B220 (clone RA3-6B2), F4/80 (Serotec), IL-1 β (R&D Systems) and TNF (PharMingen). We used secondary antibodies coupled to Alexa 488 (Molecular Probes). Sections were counterstained with phalloidin labelled with tetramethyl rhodamine B isothiocyanate (Sigma) and with ToPro III (Molecular Probes) to visualize tissue structure and nuclei. We performed TUNEL staining using the Apoptosis Detection System (Promega). Fluorescent staining was analysed using a Leica TCS upright confocal laser-scanning microscope at excitation wavelengths of 488, 543 and 633 nm.

Keratinocyte proliferation, differentiation and apoptosis

Primary epidermal keratinocytes were isolated from the skin of newborn mice as described²⁷ and cultured in FAD medium containing 50 μ M calcium ions in the absence or presence of mitomycin C-treated 3T3 fibroblasts as feeders. For induction of differentiation, keratinocytes were transferred to FAD medium containing 1.8 mM calcium ions and methylcellulose²⁸ and cultured in suspension for 24 h. Cells were recovered, washed, attached to chamber slides (Invitrogen), fixed with methanol and stained for differentiation markers. For the detection of TNF-induced apoptosis, cells were treated with 30 ng ml⁻¹ TNF for 18 h and were analysed by TUNEL staining. Images were digitized and positive cells were counted.

In situ hybridization and RNA expression analysis

In situ hybridization experiments were performed as described previously²⁹. Mouse cDNA probes were provided by J. M. Farber (Mig), T. A. Hamilton (IP-10), J. B. Smith and H. R. Herschman (MIP-2 and LIX), T. Yoshimura (MCP-1) and A. Orlofsky (C10). Briefly, paraffin-embedded tissue was cut, deparaffinized, rehydrated and acetylated. Sections were then overlaid with the hybridization solution containing ³⁵S-labelled antisense or, for control, sense probes. Slides were then dipped into Kodak NTB-2 solution, exposed for autoradiography, counterstained and microscopically evaluated. We performed RNase protection assays as described³⁰. For polymerase chain reaction with reverse transcription (RT-PCR) analysis the following primers were used: mouse IL-1 β , 5'-CTGAAGCAGCTAT GGCAACT-3' and 5'-GGATGCTCTCATCTGGACAG-3'; mouse actin, 5'-TAAACGCG AGCTCAGTAACAGTCCG-3' and 5'-TGGAATCCTGTGGCATCCATGAAAC-3'.

Received 15 February; accepted 11 April 2002; doi:10.1038/nature00820.

1. Karin, M. & Ben-Neriah, Y. Phosphorylation meets ubiquitination: the control of NF- κ B activity. *Annu. Rev. Immunol.* **18**, 621–663 (2000).

- Israel, A. The IKK complex: an integrator of all signals that activate NF- κ B? *Trends Cell Biol.* **10**, 129–133 (2000).
- Li, Q., Van Antwerp, D., Mercurio, F., Lee, K. F. & Verma, I. M. Severe liver degeneration in mice lacking the I κ B kinase 2 gene. *Science* **284**, 321–325 (1999).
- Tanaka, M. *et al.* Embryonic lethality, liver degeneration, and impaired NF- κ B activation in IKK- β -deficient mice. *Immunity* **10**, 421–429 (1999).
- Li, Z. W. *et al.* The IKK β subunit of I κ B kinase (IKK) is essential for nuclear factor κ B activation and prevention of apoptosis. *J. Exp. Med.* **189**, 1839–1845 (1999).
- Rudolph, D. *et al.* Severe liver degeneration and lack of NF- κ B activation in NEMO/IKK γ -deficient mice. *Genes Dev.* **14**, 854–862 (2000).
- Schmidt-Suppran, M. *et al.* NEMO/IKK γ -deficient mice model incontinentia pigmenti. *Mol. Cell* **5**, 981–992 (2000).
- Makris, C. *et al.* Female mice heterozygous for IKK γ /NEMO deficiencies develop a dermatopathy similar to the human X-linked disorder incontinentia pigmenti. *Mol. Cell* **5**, 969–979 (2000).
- Takeda, K. *et al.* Limb and skin abnormalities in mice lacking IKK α . *Science* **284**, 313–316 (1999).
- Li, Q. *et al.* IKK1-deficient mice exhibit abnormal development of skin and skeleton. *Genes Dev.* **13**, 1322–1328 (1999).
- Hu, Y. *et al.* Abnormal morphogenesis but intact IKK activation in mice lacking the IKK α subunit of I κ B kinase. *Science* **284**, 316–320 (1999).
- Hu, Y. *et al.* IKK α controls formation of the epidermis independently of NF- κ B. *Nature* **410**, 710–714 (2001).
- Seitz, C. S., Lin, Q., Deng, H. & Khavari, P. A. Alterations in NF- κ B function in transgenic epithelial tissue demonstrate a growth inhibitory role for NF- κ B. *Proc. Natl Acad. Sci. USA* **95**, 2307–2312 (1998).
- Seitz, C. S., Deng, H., Hinata, K., Lin, Q. & Khavari, P. A. Nuclear factor κ B subunits induce epithelial cell growth arrest. *Cancer Res.* **60**, 4085–4092 (2000).
- Beg, A. A., Sha, W. C., Bronson, R. T. & Baltimore, D. Constitutive NF- κ B activation, enhanced granulopoiesis, and neonatal lethality in I κ B-deficient mice. *Genes Dev.* **9**, 2736–2746 (1995).
- Klement, J. F. *et al.* I κ B α deficiency results in a sustained NF- κ B response and severe widespread dermatitis in mice. *Mol. Cell Biol.* **16**, 2341–2349 (1996).
- Barton, D., HogenEsch, H. & Weih, F. Mice lacking the transcription factor RelB develop T cell-dependent skin lesions similar to human atopic dermatitis. *Eur. J. Immunol.* **30**, 2323–2332 (2000).
- The International Incontinentia Pigmenti (IP) Consortium. Genomic rearrangement in NEMO impairs NF- κ B activation and is a cause of incontinentia pigmenti. *Nature* **405**, 466–472 (2000).
- Fuchs, E. & Raghavan, S. Getting under the skin of epidermal morphogenesis. *Nature Rev. Genet.* **3**, 199–209 (2002).
- Schwenk, F., Baron, U. & Rajewsky, K. A cre-transgenic mouse strain for the ubiquitous deletion of loxP-flanked gene segments including deletion in germ cells. *Nucleic Acids Res.* **23**, 5080–5081 (1995).
- Mao, X., Fujiwara, Y. & Orkin, S. H. Improved reporter strain for monitoring Cre recombinase-mediated DNA excisions in mice. *Proc. Natl Acad. Sci. USA* **96**, 5037–5042 (1999).
- Mombaerts, P. *et al.* Mutations in T-cell antigen receptor genes α and β block thymocyte development at different stages. *Nature* **360**, 225–231 (1992).
- Pfeffer, K. *et al.* Mice deficient for the 55 kD tumour necrosis factor receptor are resistant to endotoxic shock, yet succumb to *L. monocytogenes* infection. *Cell* **73**, 457–467 (1993).
- Szabowski, A. *et al.* c-Jun and JunB antagonistically control cytokine-regulated mesenchymal-epidermal interaction in skin. *Cell* **103**, 745–755 (2000).
- Kontgen, F., Suss, G., Stewart, C., Steinmetz, M. & Bluethmann, H. Targeted disruption of the MHC class II Aa gene in C57BL/6 mice. *Int. Immunol.* **5**, 957–964 (1993).
- Yamaoka, S. *et al.* Complementation cloning of NEMO, a component of the I κ B kinase complex essential for NF- κ B activation. *Cell* **93**, 1231–1240 (1998).
- Roper, E., Weinberg, W., Watt, F. M. & Land, H. p19ARF-independent induction of p53 and cell cycle arrest by Raf in murine keratinocytes. *EMBO Rep.* **2**, 145–150 (2001).
- Romero, M. R., Carroll, J. M. & Watt, F. M. Analysis of cultured keratinocytes from a transgenic mouse model of psoriasis: effects of suprabasal integrin expression on keratinocyte adhesion, proliferation and terminal differentiation. *Exp. Dermatol.* **8**, 53–67 (1999).
- Goebeler, M. *et al.* Differential and sequential expression of multiple chemokines during elicitation of allergic contact hypersensitivity. *Am. J. Pathol.* **158**, 431–440 (2001).
- Werner, S. *et al.* Targeted expression of a dominant-negative FGF receptor mutant in the epidermis of transgenic mice reveals a role of FGF in keratinocyte organization and differentiation. *EMBO J.* **12**, 2635–2643 (1993).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements

We thank B. Hampel, A. Egert, A. Leinhaas, R. Pofahl, A. Arora, R. Knaup, C. Bessia and the members of the laboratory for skin histopathology of the University of Cologne for technical support. We also thank J. Peschon, G. Kollias and S. Werner for critical reading of the manuscript, G. Mahrle for discussions and F. M. Watt and T. Magin for providing antibodies. M.P. was supported by fellowships from EMBO and the Leukemia and Lymphoma Society; I.H. received grants from the German Ministry of Education and Research and from the Köln Fortune Program. This work was supported by grants from the Cologne Centre for Molecular Medicine (ZMMK) to W. Muller and K.R., from Ligue contre le Cancer (Equipe Labellisée) to A.I., and from the Körber Foundation, the European Union and the Deutsche Forschungsgemeinschaft to K.R.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.P. (e-mail: pasparakis@embl-monterotondo.it).

Mice deficient in the Rac activator Tiam1 are resistant to Ras-induced skin tumours

Angeliki Malliri*, Rob A. van der Kammen*, Kristopher Clark*, Maarten van der Valk*, Frits Michiels† and John G. Collard*

* The Netherlands Cancer Institute, Division of Cell Biology, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Proteins of the Rho family control signalling pathways that regulate the actin cytoskeleton and gene transcription^{1,2}. *In vitro* studies have implicated Rho-like GTP-hydrolysing enzymes (GTPases) in cell migration^{3–5}, cell-cycle progression^{6,7}, and Ras-induced focus formation^{8,9}, suggesting a role for these GTPases in the formation and progression of tumours *in vivo*. To study this, we have generated mice lacking the Rac-specific activator Tiam1^{10–12}, a T-lymphoma invasion and metastasis inducing protein. Here we show that such Tiam1^{−/−} mice are resistant to the development of Ras-induced skin tumours initiated with 7,12-dimethylbenzanthracene and promoted with 12-O-tetradecanoylphorbol-13-acetate. Moreover, the few tumours produced in Tiam1^{−/−} mice grew much slower than did tumours in wild-type mice. Tiam1-deficient primary embryonic fibroblasts were also resistant to Ras^{V12}-induced focus formation. Analysis of Tiam1 heterozygotes indicated that both tumour initiation and promotion were dependent on the Tiam1 gene dose. Tiam1 deficiency was associated with increased apoptosis during initiation, and with impeded proliferation during promotion. Although the number of tumours in Tiam1^{−/−} mice was small, a greater proportion progressed to malignancy, suggesting that Tiam1 deficiency promotes malignant conversion. Our studies identify the Rac activator Tiam1 as a critical regulator of different aspects of Ras-induced tumour formation.

Tiam1-deficient mice were generated by homologous recombination. Genotyping revealed the successful generation of heterozygous (Tiam1^{+/-}) and homozygous Tiam1^{−/−} mice (Supplementary Information). Western blot analysis confirmed the absence of Tiam1 protein in the cerebellum of Tiam1^{−/−} mice, whereas Tiam1^{+/-} mice showed an intermediate expression level of Tiam1 (Fig. 1a). Consistent with the role of Tiam1 as a guanine nucleotide exchange factor (GEF) for Rac, reduced Rac activity was detected in Tiam1-deficient embryonic stem cells, and in primary keratinocytes derived from Tiam1^{−/−} mice (Fig. 1b, c). Although Tiam1 is expressed during embryonic development¹³, Tiam1^{−/−} mice develop, grow and reproduce normally. Compensation by other GEFs capable of activating Rac—such as Sos, Vav, Pix, Trio and Tiam2—may take part in the suppression of a developmental phenotype.

Tiam1 is widely expressed, with highest levels detected in brain and testis^{10,13}. In mouse skin, Tiam1 is present in basal and suprabasal keratinocytes of the interfollicular epidermis and in hair follicles, where it is predominantly expressed in the infundibular portion (Fig. 1d). Tiam1 was not detected in skin derived from Tiam1^{−/−} mice, consistent with the reduced basal Rac activity found in primary keratinocytes isolated from Tiam1^{−/−} mice (Fig. 1c).

Skin tumours were initiated in wild-type (Tiam1^{+/+}) and Tiam1^{−/−} littermates by application of a two-stage chemical carcinogenesis protocol^{14,15}. This protocol entails tumour initiation in epidermal keratinocytes by a single treatment with the carcinogen 7,12-dimethylbenzanthracene (DMBA), which invariably induces oncogenic activation of the c-Ha-Ras gene¹⁶. Subsequent repeated

treatments with the tumour promoter 12-O-tetradecanoylphorbol-13-acetate (TPA) result in the outgrowth and progression of initiated cells. Benign tumours (largely papillomas) appeared in Tiam1^{+/+} mice within 7–10 weeks from DMBA initiation. Tiam1^{−/−} mice displayed a notable resistance to tumour formation (Figs 1e and 2a, b). Papillomas developed only after a latency period of 10 weeks, and their number was on average tenfold reduced (Fig. 2a). The resistance to tumour development was also reflected in the incidence of mice bearing tumours (Fig. 2b). Immunohistochemical staining revealed high levels of Tiam1 protein expression (comparable to normal and hyperplastic epidermis) in papillomas of Tiam1^{+/+} mice (Figs 1f and 3b, and Supplementary Information). This resembled the LacZ expression driven by the Tiam1 promoter in papillomas from Tiam1^{−/−} mice (Fig. 1f).

Tiam1 deficiency affected not only tumour incidence and burden but also tumour growth. Tumours that developed in Tiam1^{−/−}

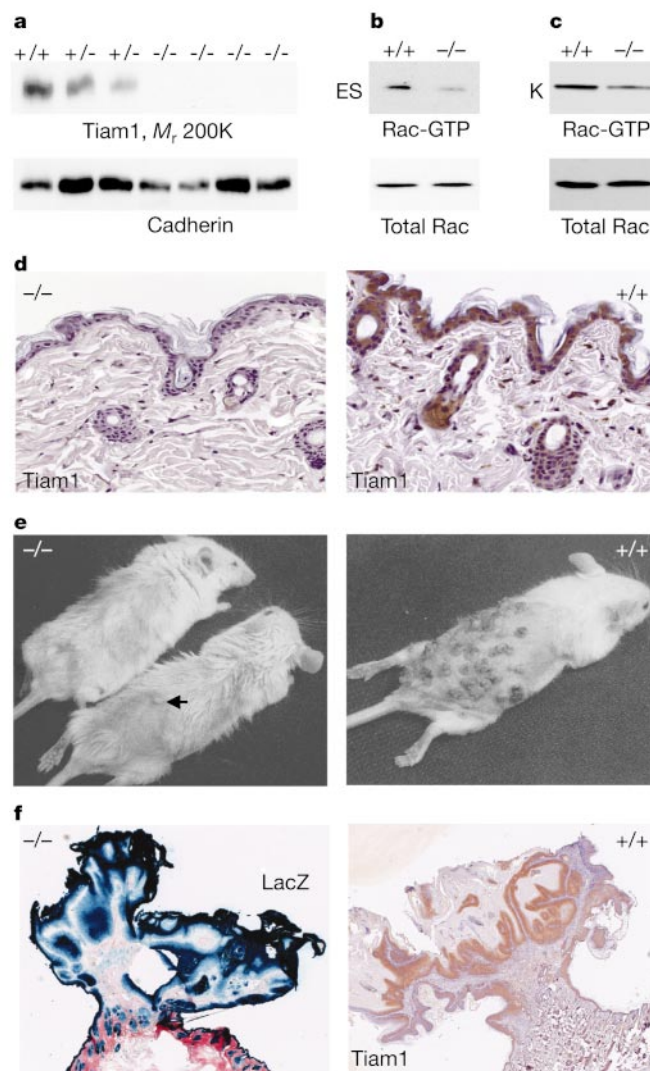


Figure 1 Analysis of Tiam1^{−/−} mice and the role of Tiam1 in papilloma development. **a**, Cerebellum protein extracts from Tiam1^{+/+}, Tiam1^{+/-} and Tiam1^{−/−} mice probed with a Tiam1-specific (top) and pan-cadherin-specific (bottom) antibody. *M_r* 200K, relative molecular mass of 200,000. **b**, **c**, Rac-GTP levels in Tiam1^{−/−} and Tiam1^{+/+} embryonic stem (ES) cells and primary keratinocytes (K). **d**, Tiam1 expression in the skin from Tiam1^{−/−} and Tiam1^{+/+} mice. **e**, Papilloma development in Tiam1^{−/−} and Tiam1^{+/+} mice after DMBA/TPA treatment (18 weeks). Tiam1^{−/−} mice developed no or very few papillomas (arrow). **f**, Papilloma from a Tiam1^{−/−} and Tiam1^{+/+} mouse stained for LacZ activity and Tiam1 expression, respectively. LacZ expression is driven by the Tiam1 promoter in Tiam1^{−/−} mice.

† Present address: Galapagos Genomics, Archimedesweg 4, 2333 CN Leiden, The Netherlands.

mice were significantly smaller than tumours appearing in *Tiam1*^{+/+} mice (Fig. 2c). Consistent with the *in vivo* data, focus formation induced by *Ras*^{V12} and *Myc* in primary mouse embryonic fibroblasts derived from *Tiam1*^{-/-} mice was strongly reduced when compared to wild-type mouse embryonic fibroblasts (Fig. 2e), supporting a direct role for Tiam1 in Ras-induced transformation of cells. The frequency and growth of tumours in heterozygous *Tiam1*^{+/-} mice was intermediate between that of *Tiam1*^{+/+} and *Tiam1*^{-/-} mice (Fig. 2c, d). Therefore, we conclude that *Tiam1* gene dosage critically affects the efficiency of Ras-induced tumour initiation, as determined by tumour numbers, as well as TPA-induced tumour promotion, as reflected in tumour growth.

To further discriminate between the effects of Tiam1 deficiency on DMBA-induced tumour initiation versus TPA-induced tumour promotion, we also treated groups of mice repeatedly with DMBA alone, rather than with DMBA and TPA. This complete carcinogenesis protocol leads directly to the formation of predominantly squamous cell carcinomas (SCC). The average number of tumours produced in *Tiam1*^{-/-} mice treated with DMBA alone was significantly reduced when compared to *Tiam1*^{+/+} mice (Fig. 2f), as similarly found in animals treated with DMBA/TPA (Fig. 2a). However, the growth rate of tumours induced by DMBA alone was not different between *Tiam1*^{+/+} and *Tiam1*^{-/-} mice (Fig. 2g). Thus, tumour growth was only inhibited in *Tiam1*^{-/-} mice treated with both DMBA and TPA, and was not inhibited in mice treated with DMBA alone (compare Figs 2c and g). These data substantiate

a role for Tiam1 in DMBA-induced tumour initiation, and indicate that the effects of Tiam1 on tumour growth are due to an additional requirement for Tiam1 during TPA-induced tumour promotion.

We observed codon 61 c-Ha-Ras mutations in all the tumours that we tested from *Tiam1*^{+/+} and *Tiam1*^{-/-} mice (see Supplementary Information), indicating that Tiam1-deficiency did not affect the ability of cells to acquire c-Ha-Ras mutations. However, DMBA but not TPA treatment induced significantly more apoptosis in the basal layer of the epidermis of *Tiam1*^{-/-} mice compared to *Tiam1*^{+/+} mice (Fig. 2h). It seems that lack of Tiam1 renders cells more susceptible to DMBA-induced apoptosis, which could explain the resistance of *Tiam1*^{-/-} mice to tumour initiation. Analysis of cell proliferation in the basal layer of the hyperplastic epidermis of mice following treatment with TPA or DMBA/TPA revealed that the number of cycling cells in *Tiam1*^{-/-} mice was significantly reduced compared to *Tiam1*^{+/+} mice (Fig. 2i). This is consistent with the reduced growth rate of tumours induced by DMBA/TPA in *Tiam1*^{-/-} mice. In contrast, treatment with DMBA alone induced an equivalent number of cycling cells in the basal layer of the epidermis of *Tiam1*^{+/+} and *Tiam1*^{-/-} mice (Fig. 2i), in line with the equal growth of tumours induced by DMBA alone in mice of either genotype (Fig. 2g). These data further support a role for Tiam1 in TPA-induced tumour growth. Thus, our findings indicate that Tiam1 is required for DMBA-induced tumour initiation as well as for TPA-induced tumour promotion, probably by two independent pathways, involving apoptosis and cell proliferation respectively.

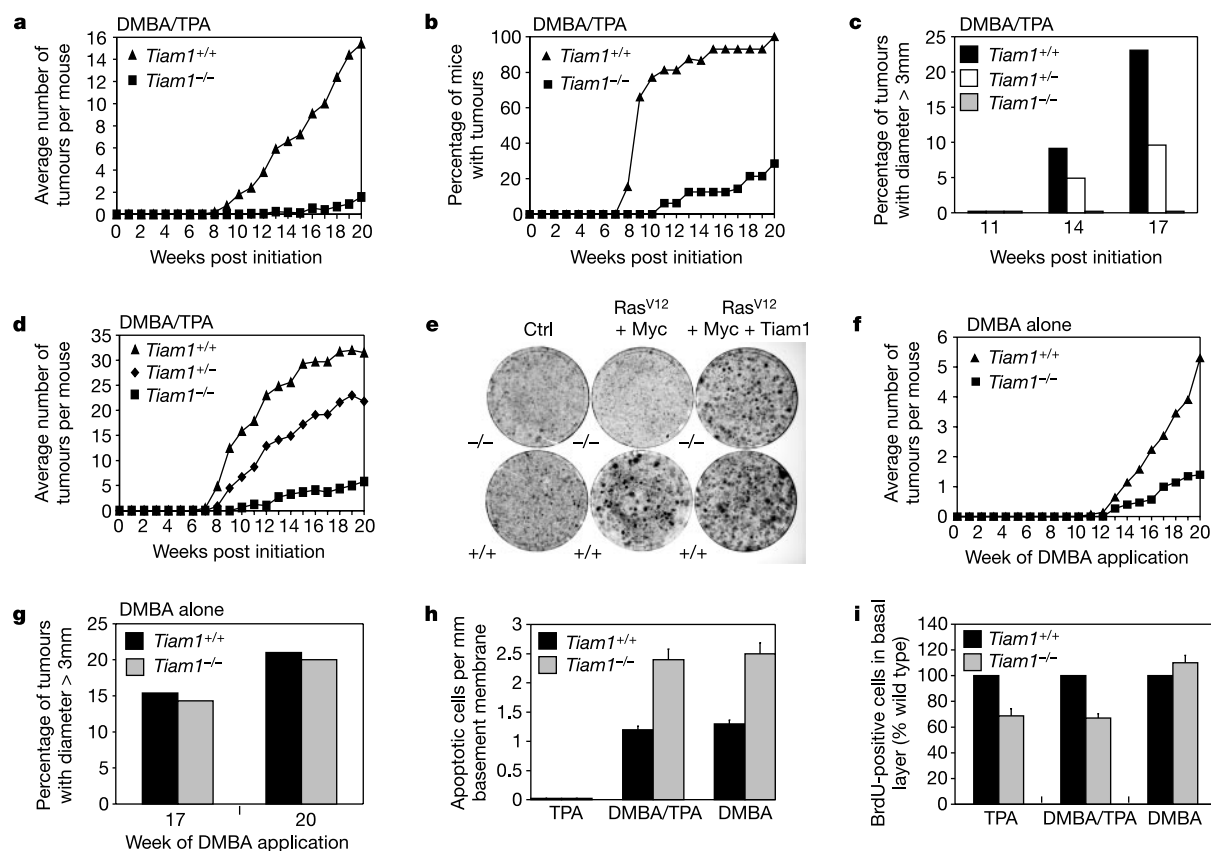


Figure 2 Skin carcinogenesis in wild-type and *Tiam1* mutant mice. **a–d**, Mice were subjected to the two-stage chemical carcinogenesis protocol using DMBA (7,12-dimethylbenzanthracene) and TPA (12-*O*-tetradecanoylphorbol-13-acetate). **a**, Average number of tumours per mouse; **b**, incidence of tumours; and **c**, growth of tumours. The difference in growth rate is statistically significant ($P < 0.001$; χ^2 test). *Tiam1*^{+/-} mice display an intermediate tumour susceptibility phenotype (**d**). **e**, Focus formation induced by *Ras*^{V12} and *Myc* or *Ras*^{V12}, *Myc* and *Tiam1* in primary embryonic fibroblasts isolated from *Tiam1*^{-/-} and *Tiam1*^{+/+} mice. **f, g**, Mice were treated with DMBA alone for 20

weeks. Average number of tumours per mouse (**f**) and growth of tumours (**g**). **h**, Apoptosis induced in the skin of mice 24 h after a single application of TPA, DMBA or DMBA/TPA. The difference in DMBA- and TPA-induced apoptosis between the two genotypes is statistically significant ($P < 0.001$; Student's *t*-test). **i**, Cell proliferation induced in the skin of mice after a single application of either TPA or DMBA, or after a single application of DMBA followed by two applications of TPA. The degree of proliferation in *Tiam1*^{+/+} mice is set at 100%. The difference in TPA or DMBA/TPA-induced cell proliferation between the two genotypes is statistically significant ($P < 0.001$; Student's *t*-test).

As *in vitro* experiments have implicated Tiam1 (refs 3, 11) and Rac^{4,5} in cell migration and cell invasion, we studied the effects of Tiam1 deficiency on tumour progression. As skin tumours rarely metastasize, tumour progression was investigated by histological analysis. DMBA/TPA-treated *Tiam1*^{+/+} mice developed a variety of skin tumours, predominantly benign papillomas of which approximately 35% ultimately progressed to malignancy. In contrast, approximately 80% of the tumours that arose in *Tiam1*^{-/-} mice acquired an invasive, malignant phenotype (Fig. 3a, Table 1a). The malignancies found in both *Tiam1*^{+/+} and *Tiam1*^{-/-} mice comprised a similar spectrum of histopathological types (Table 1a). Most malignant tumours were classified as severely dysplastic papillomas displaying carcinoma *in situ* and/or focal invasion, or were squamous cell carcinomas (SCCs). A smaller percentage of malignancies comprised spindle cell carcinomas (Table 1a). Thus, Tiam1 deficiency seems to increase the frequency of malignant conversion in animals treated with DMBA/TPA. Analysis of tumours derived from mice treated with DMBA alone further supported a role for Tiam1 in malignant progression. Carcinogenesis by repeated treatment with DMBA alone gave rise to virtually only SCCs (>90%) in both *Tiam1*^{+/+} and *Tiam1*^{-/-} mice. However, the SCCs derived from *Tiam1*^{-/-} mice showed a significantly

higher malignancy grade than tumours from wild-type animals (Table 1b and Fig. 4). Furthermore, immunohistochemical analysis of Tiam1 expression in tumours of wild-type mice revealed that benign papillomas maintained high levels of Tiam1 expression in a similar fashion to normal and hyperplastic epidermis (Figs 1f and 3b). However, Tiam1 expression was reduced in SCCs, and this reduction correlated with malignant progression and loss of differentiation of these SCCs (Figs 3b and 4, and Supplementary Information). Tiam1 could not be detected in the spindle cell carcinomas or in spindle components of lesions consisting of both spindle and papillomatous areas (Fig. 3c, and Supplementary Information).

Previously, we found that sustained oncogenic Ras-signalling in epithelial Madin–Darby canine kidney (MDCK) cells results in down regulation of Tiam1 expression, a decrease in Rac activity, and epithelial–mesenchymal transition¹⁹. It is possible therefore that increased Ras-signalling associated with advanced skin malignancies^{15,16} is responsible for the reduction or loss of Tiam1 expression in these lesions. Indeed, A5 cells^{17,18} derived from a spindle cell carcinoma display an increased amount of active, mutant Ras protein and simultaneously a loss of Tiam1 protein when compared to P1 cells^{17,18}, derived from a benign papilloma (Fig. 3d). Moreover, the loss of Tiam1 in the spindle A5 cells was accompanied by a strong reduction in basal Rac activity (Fig. 3d). These data support the histochemical analysis of the tumours, and are consistent with the downregulation of Tiam1 expression in established Ras^{V12}-transformed MDCK cells¹⁹. Thus, although Tiam1 function appears to be essential for the initiation and promotion of Ras-induced skin tumours, the histological and biochemical data suggest that a subsequent loss of Tiam1 increases the malignant conversion of benign tumours. A similar biphasic action in skin chemical carcinogenesis has been reported in *p53*-deficient mice, where the deficiency promotes malignant progression but reduces the total number of tumours²⁰.

Our data provide experimental evidence for a role of a Rac-specific activator in tumorigenesis *in vivo*. Tiam1-mediated Rac signalling is required for effective initiation of epidermal tumours by oncogenic Ras and for tumour promotion by TPA in a manner that depends on the gene dosage of *Tiam1*. These findings implicate Tiam1 as a critical component of Ras-induced tumour formation *in vivo*, and are consistent with the requirement for Rac signalling in Ras^{V12}-induced focus formation in established fibroblast cultures^{8,9}. Indeed, similarly to Rac, Tiam1 is required for Ras^{V12}-induced focus formation in primary embryonic fibroblasts. Tiam1 deficiency

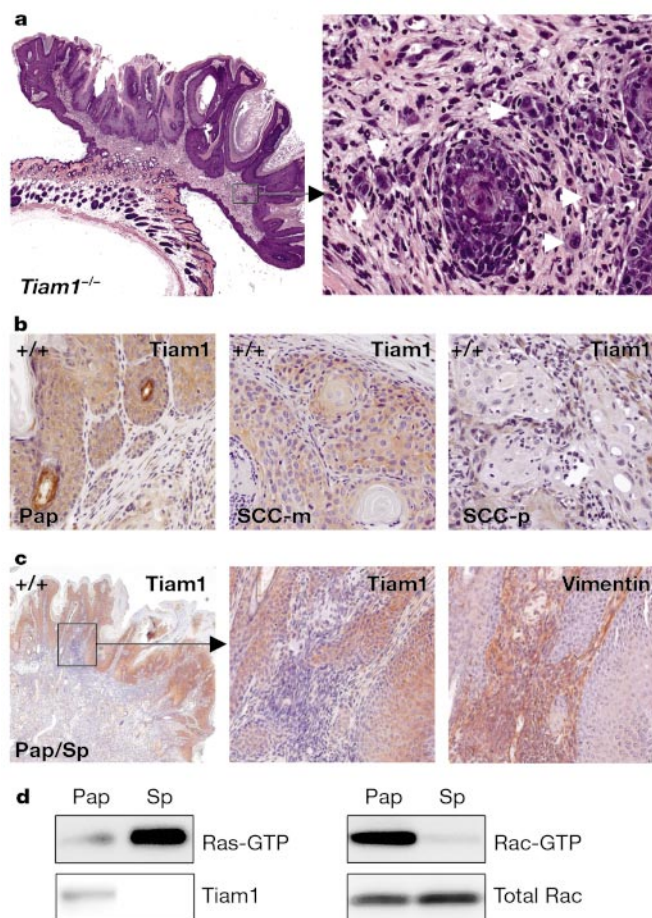


Figure 3 Tiam1 and malignant progression. **a**, A lesion from a *Tiam1*^{-/-} mouse, stained with haematoxylin and eosin. Boxed area is magnified and shows tumour cells invading the stroma (white arrowheads). **b**, Tiam1 expression in a papilloma (Pap), and in a moderately and a poorly differentiated squamous cell carcinoma (SCC-m and SCC-p) from wild-type mice (see also Supplementary Information). **c**, Complex lesion from a wild-type mouse comprising papillomatous and spindle components stained for Tiam1 (left and middle) and vimentin (right). Boxed region is magnified. Note the disappearance of Tiam1 expression in the spindle component (see also Supplementary Information). **d**, The degree of Ras- and Rac activity as well as the Tiam1 and Rac protein levels in A5 and P1 cells^{17,18} derived from a spindle cell carcinoma and a benign papilloma, respectively.

Table 1 Histological analysis of skin tumours

(a) Histological analysis of skin tumours from *Tiam1*^{+/+} and *Tiam1*^{-/-} mice subjected to the DMBA/TPA protocol*

	+/+		-/-	
	n	%	n	%
Benign lesions (papillomas and keratoacanthomas)	98	63%	4	21%
Malignant lesions	57	37%	15	79%
Complex lesions with focal invasion	(26)	(17%)	(7)	(37%)
Squamous cell carcinomas (SCC)	(21)	(14%)	(6)	(32%)
Spindle cell carcinomas	(10)	(6%)	(2)	(10%)

(b) Grading of SCCs induced by treatment with DMBA alone in *Tiam1*^{+/+} and *Tiam1*^{-/-} mice†

	+/+		-/-	
	n	%	n	%
Well differentiated SCC	15	24%	4	10%
Moderately differentiated SCC	36	57%	19	49%
Poorly differentiated SCC	12	19%	16	41%

*The difference in the incidence of malignant tumours between *Tiam1*^{+/+} and *Tiam1*^{-/-} mice is statistically significant ($P < 0.001$) using the χ^2 test. For examples, see Fig. 3b, c, and Supplementary Information for these figures.

†The difference in malignancy grade between *Tiam1*^{+/+} and *Tiam1*^{-/-} mice is statistically significant ($P < 0.05$) using the χ^2 test. For examples, see Fig. 4.

impairs the survival and proliferative capacity of cells initiated with DMBA and promoted with TPA. These functions of Tiam1 are in agreement with *in vitro* studies that implicate Rac, through activation of NF- κ B, in the suppression of apoptosis elicited by oncogenic Ras²¹. Rac also has a role in cell cycle progression⁶ and synergizes with the Raf/MAPK pathway to induce optimal cyclin D1 expression (reviewed in ref. 2), which regulates the G1–S phase transition. Similarly to Tiam1, cyclin D1 is essential for oncogenic Ras-induced skin tumours in mice²². Ras could activate Tiam1 through PI(3) kinase, a direct target of Ras, as Tiam1-mediated Rac activation is dependent on PI(3) kinase activity²³. Alternatively, Tiam1 could be activated through direct binding to active GTP-bound Ras via the Ras binding domain in the Tiam1 protein, as has recently been found (C. Der, personal communication).

With respect to tumour progression, Tiam1 deficiency seems to favour malignant progression of epithelial tumours *in vivo*. It can not be excluded that Tiam1-deficient cells require more mutations than their wild-type counterparts to establish benign papillomas, with, as a consequence, an increased potential for malignant progression. However, given the reduction and loss of Tiam1 expression in the advanced tumours in wild-type mice, it is conceivable that the lack of Tiam1 in the *Tiam1*^{-/-} mice has a causal role in promoting tumour progression. *In vitro* studies have implicated Rac in cell migration and invasion of a large variety of tumour cell lines^{4,5,11,24}. Investigations in epithelial MDCK cells revealed that Tiam1 and activated Rac promote the formation and maintenance of E-cadherin-mediated adhesions on most extracellular matrix substrates, and thereby inhibit the migration of epithelial cells^{3,24–26}. Such decreased adhesion has been correlated with the progression of epithelial tumours^{27,28}. Tiam1 deficiency or loss in epithelial skin tumours might therefore favour malignant progression and invasion by decreasing Rac activity, thus reducing the strength of E-cadherin-based adhesions. Taken together, our data illustrate an important role for Tiam1, an activator of Rac, in different aspects of Ras-induced tumorigenesis. □

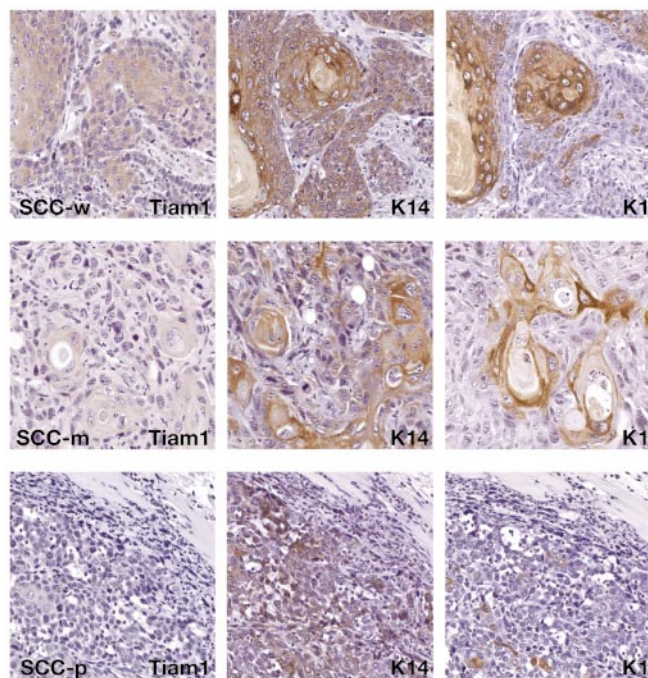


Figure 4 Expression of Tiam1, keratin 1 and keratin 14 in squamous cell carcinomas. Tiam1, keratin 1 (K1) and keratin 14 (K14) expression is shown in a well, a moderately, and a poorly differentiated squamous cell carcinoma (SCC-w, SCC-m, SCC-p) induced by the DMBA-alone protocol in wild-type mice.

Methods

Generation of *Tiam1*^{-/-} mice and genotyping

See Supplementary Information.

Carcinogenesis protocols and analysis of tumours

For two-stage chemical carcinogenesis, the backs of 8-week-old mice were shaved and treated with a single application of DMBA (25 μ g in 200 μ l acetone; Sigma) followed by biweekly applications of TPA (200 μ l of 10⁻⁴ M solution in acetone; Sigma) for 20 weeks. Cohorts of *Tiam1*^{+/+}, *Tiam1*^{+/-} and *Tiam1*^{-/-} mice of equal sex ratio were used, consisting of littermates generated from intercrosses of eighth-generation *Tiam1*^{+/-} mice on an FVB background. For complete carcinogenesis, mice were treated biweekly with 5 μ g DMBA alone in 200 μ l acetone for 20 weeks. Mice were visually examined weekly and were killed if moribund, if any individual tumour reached a diameter of 1 cm, or at the termination of the experiments. Tumours and organs from dead animals were snap-frozen in liquid nitrogen, fixed in formalin and stained with haematoxylin and eosin.

Immunohistochemical staining

This was performed on paraffin-embedded tissue sections (4 μ m) using a Tiam1-specific antibody¹⁰, a vimentin-specific antibody (Innogenex), or keratin 1 or 14 specific antibodies (BabCO).

Apoptosis in skin

TPA (10⁻⁴ M) or DMBA (25 μ g) or both were applied to the backs of five mice of each genotype; the mice were killed 24 h later. Apoptotic cells in skin biopsy specimens were detected by TUNEL (TdT-mediated dUTP nick end labelling) assay using the ApopTag Plus Peroxidase *in situ* Apoptosis Detection Kit (Intergen). The number of TUNEL-positive epidermal keratinocytes in the basal layer of interfollicular regions from several (>30) independent fields was evaluated per millimetre of basement membrane by microscopy.

Cell proliferation in skin

Groups of mice ($n = 5$) were treated with TPA alone (10⁻⁴ M) or DMBA alone (25 μ g) for 24 h, or were treated with DMBA and subsequently treated with 2 doses of TPA over 7 d. At the end of the experiment, mice were injected intraperitoneally with BrdU (Sigma) at 50 mg kg⁻¹, and killed 2 h later. Using an anti-BrdU antibody (DAKO), the number of BrdU-positive epidermal keratinocytes in the basal layer of interfollicular regions from independent fields was evaluated per 100 basal keratinocytes.

Pull down and focus formation assays

The amounts of active GTP-bound Rac²³ and GTP-bound Ras²⁹ in cells were determined using pull down assays. Focus formation was investigated by infecting primary embryonic fibroblasts isolated from *Tiam1*^{+/+} and *Tiam1*^{-/-} mice with retroviruses coding for Ha-Ras^{V12} and c-Myc, Ha-Ras^{V12}, c-Myc and Tiam1, or empty virus only as control. Cells were allowed to grow for 14 d, fixed, and stained with crystal violet.

Received 21 February; accepted 24 April 2002; doi:10.1038/nature00848.

- Hall, A. Rho GTPases and the actin cytoskeleton. *Science* **279**, 509–514 (1998).
- Bar-Sagi, D. & Hall, A. Ras and Rho GTPases: a family reunion. *Cell* **103**, 227–238 (2000).
- Hordijk, P. L. *et al.* Inhibition of invasion of epithelial cells by Tiam1-Rac signalling. *Science* **278**, 1464–1466 (1997).
- Keely, P. J., Westwick, J. K., Whitehead, I. P., Der, C. J. & Parise, L. V. Cdc42 and Rac1 induce integrin-mediated cell motility and invasiveness through PI(3)K. *Nature* **390**, 632–636 (1997).
- Shaw, L. M., Rabinovitz, L., Wang, H. H. F., Toker, A. & Mercurio, A. M. Activation of phosphoinositide 3-OH kinase by the α 6 β 4 integrin promotes carcinoma invasion. *Cell* **91**, 949–960 (1997).
- Olson, M. F., Ashworth, A. & Hall, A. An essential role for Rho, Rac, and Cdc42 GTPases in cell cycle progression through G1. *Science* **269**, 1270–1272 (1995).
- Olson, M. F., Paterson, H. F. & Marshall, C. J. Signals from Ras and Rho GTPases interact to regulate expression of p21Waf1/Cip1. *Nature* **394**, 295–299 (1998).
- Qiu, R. G., Chen, J., Kirm, D., McCormick, F. & Symons, M. An essential role for Rac in Ras transformation. *Nature* **374**, 457–459 (1995).
- Khosravi-Far, R., Solski, P. A., Clark, G. J., Kinch, M. S. & Der, C. J. Activation of Rac and Rho, and mitogen activated protein kinases, are required for Ras transformation. *Mol. Cell. Biol.* **15**, 6443–6453 (1995).
- Habets, G. G. M. *et al.* Identification of an invasion-inducing gene, Tiam1, that encodes a protein with homology to GDP-GTP exchangers for Rho-like proteins. *Cell* **77**, 537–549 (1994).
- Michiels, F., Habets, G., Stam, J., van der Kammen, R. & Collard, J. A role for Rac in Tiam1-induced membrane ruffling and invasion. *Nature* **375**, 338–340 (1995).
- Worthylake, D., Rossman, K. & Sondek, J. Crystal structure of Rac1 in complex with the guanine nucleotide exchange region of Tiam1. *Nature* **408**, 682–688 (2000).
- Ehler, E., van Leeuwen, E., Collard, J. & Salinas, P. Expression of Tiam-1 in the developing brain suggests a role for the Tiam-1-Rac signalling pathway in cell migration and neurite outgrowth. *Mol. Cell Neurosci.* **9**, 1–12 (1997).
- Yuspa, S. H. The pathogenesis of squamous cell cancer: lessons learned from studies of skin carcinogenesis. *J. Dermatol. Sci.* **17**, 1–7 (1998).
- Frame, S. & Balmain, A. Integration of positive and negative growth signals during ras pathway activation *in vivo*. *Curr. Opin. Genet. Dev.* **10**, 106–113 (2000).
- Quintanilla, M., Brown, K., Ramsden, M. & Balmain, A. Carcinogen-specific mutation and amplification of Ha-ras during mouse skin carcinogenesis. *Nature* **322**, 78–80 (1986).
- Zoumpouris, V. *et al.* High levels of phosphorylated c-Jun, Fra-1, Fra-2 and ATF-2 proteins correlate with malignant phenotypes in the multistage mouse skin carcinogenesis model. *Oncogene* **19**, 4011–4021 (2000).
- Buchmann, A., Ruggeri, B., Klein-Szanto, A. J. P. & Balmain, A. Progression of squamous carcinoma cells to spindle carcinomas of mouse skin is associated with an imbalance of H-Ras alleles on

- chromosome 7. *Cancer Res.* **51**, 4097–4101 (1991).
19. Zondag, G. C. M. *et al.* Oncogenic Ras downregulates Rac activity, which leads to increased Rho activity and epithelial-mesenchymal transition. *J. Cell Biol.* **149**, 775–781 (2000).
 20. Kemp, C. J., Donehower, L. A., Bradley, A. & Balmain, A. Reduction of p53 gene dosage does not increase initiation or promotion but enhances malignant progression of chemically induced skin tumors. *Cell* **74**, 813–822 (1993).
 21. Joneson, T. & Bar-Sagi, D. Suppression of Ras-induced apoptosis by the Rac GTPase. *Mol. Cell Biol.* **19**, 5892–5901 (1999).
 22. Robles, A. I. *et al.* Reduced skin tumour development in cyclin D1-deficient mice highlights the oncogenic ras pathway in vivo. *Genes Dev.* **12**, 2469–2474 (1998).
 23. Sander, E. E. *et al.* Matrix-dependent Tiam1/Rac signalling in epithelial cells promotes either cell-cell adhesion or cell migration and is regulated by phosphatidylinositol 3-kinase. *J. Cell Biol.* **143**, 1385–1398 (1998).
 24. Sander, E. E., ten Klooster, J. P., van Delft, S., Van der Kammen, R. A. & Collard, J. G. Rac downregulates Rho activity: reciprocal balance between both GTPases determines cellular morphology and migratory behaviour. *J. Cell Biol.* **147**, 1009–1022 (1999).
 25. Braga, V. M. M., Machesky, L. M., Hall, A. & Hotchin, N. A. The small GTPases rho and rac are required for the establishment of cadherin-dependent cell-cell contacts. *J. Cell Biol.* **137**, 1421–1431 (1997).
 26. Takaishi, K., Sasaki, T., Kotani, H., Nishioka, H. & Takai, Y. Regulation of cell-cell adhesion by Rac and Rho small G proteins in MDCK cells. *J. Cell Biol.* **139**, 1047–1059 (1997).
 27. Bracke, M. E., Van Roy, F. M. & Mareel, M. M. The e-cadherin/catenin complex in invasion and metastasis. *Curr. Top. Microbiol. Immunol.* **213**, 123–161 (1996).
 28. Perl, A. K., Wilgenbus, P., Dahl, U., Semb, H. & Christofori, G. A causal role for E-cadherin in the transition from adenoma to carcinoma. *Nature* **392**, 190–193 (1998).
 29. de Rooij, J. & Bos, J. L. Minimal Ras-binding domain of Raf1 can be used as an activation-specific probe for Ras. *Oncogene* **14**, 623–625 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements

We thank J.-Y. Song and the Animal Pathology department of the Netherlands Cancer Institute for histological analysis; D. Stuart, P. Demant and A. Hurlstone for advice; P. Krimpenfort and T. Schrauwers for experimental support; V. Zoumbourlis and A. Balmain for cell lines; and the animal facility of the Netherlands Cancer Institute for providing animal care. Part of this work was supported by the Dutch Cancer Society, the European Community, and the Association for International Cancer Research.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.G.C. (e-mail: j.collard@nki.nl).

Efp targets 14-3-3 σ for proteolysis and promotes breast tumour growth

Tomohiko Urano*†, Tomoyuki Saito*†‡, Tohru Tsukui§, Masayo Fujita*, Takayuki Hosoi*, Masami Muramatsu§, Yasuyoshi Ouchi* & Satoshi Inoue*§||

* Department of Geriatric Medicine, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan

§ Research Center for Genomic Medicine and Department of Molecular Biology, Saitama Medical School, 1397-1 Yamane, Hidaka-shi, Saitama 350-1241, Japan || CREST, Japan Science and Technology Corporation, Kawaguchi, Saitama, 332-0012, Japan

† These authors contributed equally to this work

Oestrogen exerts its influence on target organs through activating oestrogen receptors (ERs) and regulating downstream genes by means of their oestrogen-responsive elements. Efp, a target gene product of ER α ^{1–3}, is a member of the RING-finger B-box coiled-coil (RBCC) motif family⁴. Efp is predominantly expressed

in various female organs² as well as in breast cancers⁵, and is thought to be essential for oestrogen-dependent cell proliferation and organ development—Efp-disrupted mice display underdeveloped uteri and reduced oestrogen responsiveness⁶. Here we show that Efp is a RING-finger-dependent ubiquitin ligase (E3) that targets proteolysis of 14-3-3 σ , a negative cell cycle regulator that causes G2 arrest⁷. We demonstrate that tumour growth of breast cancer MCF7 cells implanted in female athymic mice is reduced by treatment with antisense Efp oligonucleotide. Efp-overexpressing MCF7 cells in ovariectomized athymic mice generate tumours in the absence of oestrogen. Loss of Efp function in mouse embryonic fibroblasts results in an accumulation of 14-3-3 σ , which is responsible for reduced cell growth. These data provide an insight into the cell-cycle machinery and tumorigenesis of breast cancer by identifying 14-3-3 σ as a target for proteolysis by Efp, leading to cell proliferation.

We have used MCF7 cells as a model system to explore the contribution of Efp to growth of breast tumours. MCF7 cells were implanted into intact female athymic mice followed by ovariectomy or administration of sense/antisense Efp oligonucleotides (Fig. 1a–d). Levels of Efp were markedly reduced in tumours obtained from ovariectomized or antisense Efp-treated mice compared with controls, suggesting that both treatments impaired Efp expression (Fig. 1a). Tumour volume did not increase in ovariectomized mice, consistent with the oestrogen-dependent nature of MCF7 cells (Fig. 1d). Tumour growth was observed in control mice (Fig. 1c, d); however, a dose-dependent inhibition of tumour growth was observed in mice treated with antisense Efp (Fig. 1b–d). We infer that tumour growth is modulated by Efp expression, implicating Efp as an oncogenic factor in breast cancers.

To test this hypothesis, we generated MCF7 cells that stably expressed Efp (Efp-MCF7 cells) and investigated whether Efp modulates progression of the cell cycle. Endogenous levels of p21^{Cip1} and 14-3-3 σ were reduced in Efp-MCF7 cells as compared with control MCF7 cells (vector-MCF7 cells) (Fig. 2a). These proteins are negative regulators of cell cycle progression and are important for G1 and G2 arrest after DNA damage^{7,8}, suggesting that Efp may regulate cell cycle checkpoints (Fig. 2a). This was confirmed by cell cycle analysis using flow cytometry, as a high percentage of Efp-MCF7 cells were in the proliferating stage, particularly in S phase, whereas vector-MCF7 cells were primarily in the G1 phase (Fig. 2b). To determine whether Efp overexpression influences tumour growth, we inoculated transfected MCF7 cells into intact female athymic mice. After two months, both vector-MCF7 and Efp-MCF7 cells showed tumour formation; however, the tumours were larger in Efp-MCF7 mice (data not shown). The role of Efp as a primary regulator of oestrogen-dependent signalling was verified by implanting transfected MCF7 cells into ovariectomized mice. Tumours did not grow in vector-MCF7 mice, whereas prominent tumour growth was observed in ovariectomized mice inoculated with Efp-MCF7 cells (Fig. 2c). To further confirm the role of Efp in tumour growth, transfected MCF7 cells were inoculated into intact athymic mice, followed by ovariectomy when the tumour volume reached 180 mm³ (Fig. 2d). Tumour volume in vector-MCF7 animals did not increase further, whereas tumours in Efp-MCF7 animals continued to grow in the absence of gonadal oestrogens. This indicates that elevated levels of Efp promote cell growth, indicating that Efp might directly regulate the cell cycle machinery.

As a first step towards understanding the role of Efp in cell cycle progression, we used yeast two-hybrid screening to identify Efp-interacting proteins. Notably, a single interacting clone coding for 14-3-3 σ was identified by screening 3×10^6 yeast clones transformed with a mouse embryo complementary DNA library. 14-3-3 σ was originally identified as an epithelial-specific marker, HME1, which is downregulated in a few breast and colon cancer cell lines^{7,9}. Although 14-3-3 σ seems to be an important negative cell cycle

‡ Present address: Pharmacological Research Department, Teikoku Hormone MFG Co. Ltd, 1604 Shimokusunobe, Takatsu-ku, Kawasaki, Japan.

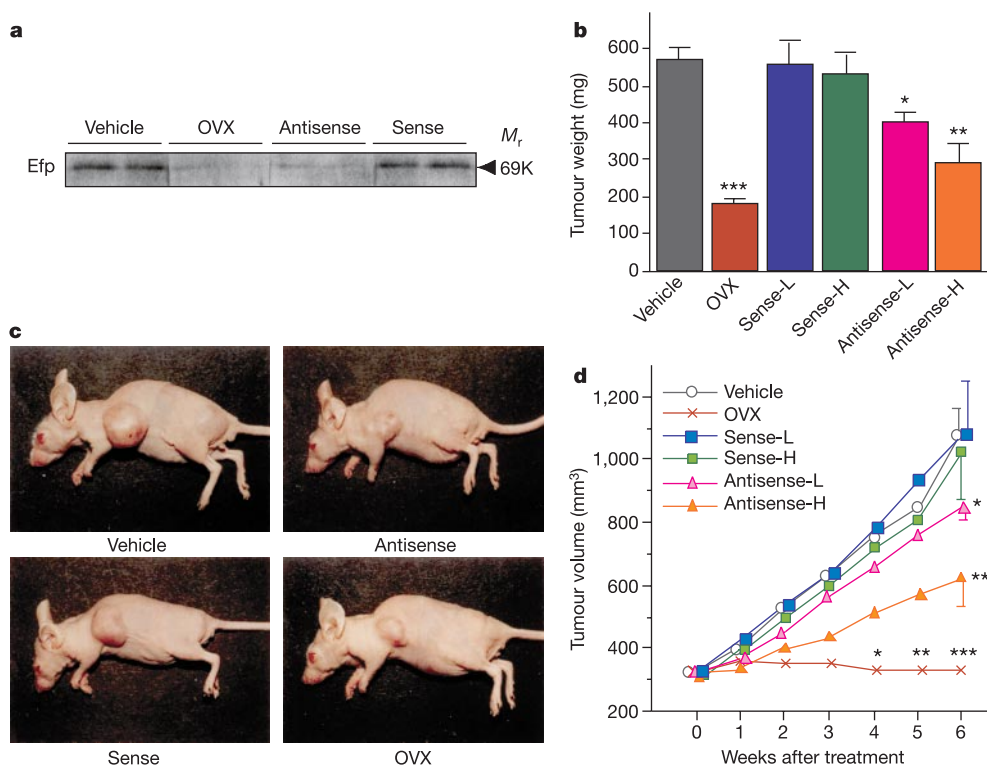


Figure 1 Inhibition of Efp expression suppresses tumour growth by MCF7 cells. Athymic female mice were inoculated with MCF7 cells then treated with an ovariectomy (OVX), or administration of vehicle (saline) or sense/antisense Efp oligonucleotides (sense-L and antisense-L, 20 μg per week; sense-H/sense and antisense-H/antisense, 100 μg per week). **a**, Efp expression in tumours obtained from mice inoculated with MCF7 cells.

b, Weight of tumours caused by MCF7 cells in nude mice. **c**, Reduced tumour size in mice 6 weeks after inoculation with antisense Efp or with OVX. **d**, Reduced tumour volume in mice treated with antisense Efp or with OVX. Asterisk, $P < 0.05$; double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$; $n = 5$.

regulator, a regulatory mechanism for 14-3-3 σ activity was previously unknown.

Next, we tested whether Efp can specifically interact with and regulate the activity of 14-3-3 σ in mammalian cells. COS7 cells were transiently transfected with Efp and 14-3-3 σ . Both proteins localized in the cytoplasm (Fig. 3a) and immunoprecipitation confirmed that Efp and 14-3-3 σ formed a complex in the cells (Fig. 3b).

In contrast, we did not detect interaction between Efp and other cell cycle checkpoints such as p21^{Cip1}, p27^{Kip1} and p57^{Kip2} (data not shown). To define 14-3-3 σ -binding domains in Efp, we constructed a series of Efp deletion mutants and analysed their binding to full-length 14-3-3 σ by immunoprecipitation (Fig. 3c). Neither amino-terminal Efp (residues 1–122) containing the RING finger domain nor carboxy-terminal Efp (residues 440–630) containing the SPRY

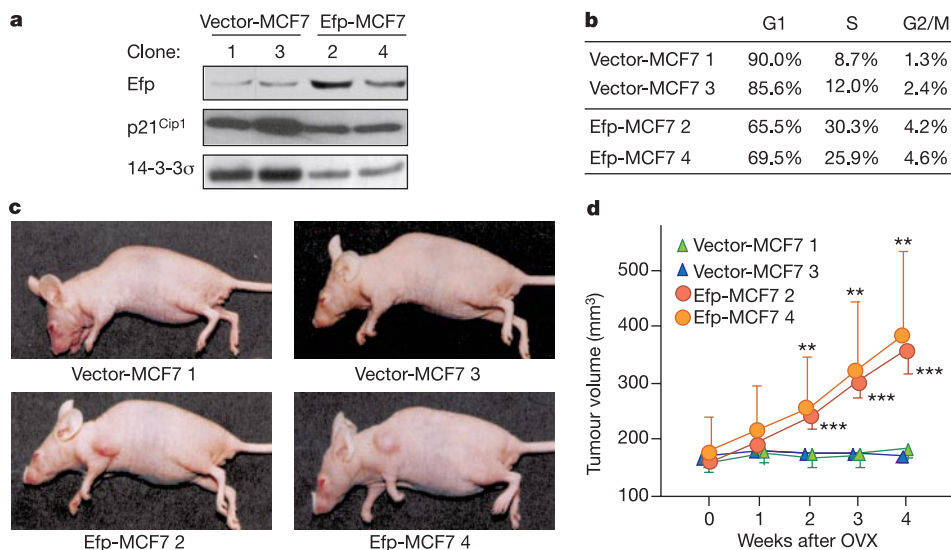


Figure 2 Efp promotes cell cycle progression and tumour growth by MCF7 cells in an oestrogen-independent manner. **a**, Efp overexpression downregulates levels of 14-3-3 σ or p21^{Cip1} in MCF7 cells. Efp-MCF7 cells stably express Efp. **b**, Efp overexpression increases the percentage of cells in the proliferation stage of the cell cycle. **c**, Hormone-

independent tumour growth is observed in ovariectomized nude mice bearing Efp-MCF7 cells. Mice were inoculated with MCF7 cells and observed after 2 months. **d**, Tumour volume of mice bearing MCF7 transfectants after ovariectomy. Mice were ovariectomized at a tumour volume of 180 mm^3 . Triple asterisk, $P < 0.001$.

domain could bind to 14-3-3 σ . However, Efp (81–450) containing the B-box coiled-coil domain was able to interact with 14-3-3 σ . The B-box coiled-coil domain often associates with a RING finger to form a larger conserved motif, RBCC⁴, and it is this motif in Efp that specifically interacts with 14-3-3 σ . It is notable that the RBCC motif family includes several oncogenic proteins such as PML¹⁰ and TIF1 (ref. 11). We next investigated whether Efp overexpression affects

the expression of 14-3-3 σ . Co-transfection of Efp and 14-3-3 σ into COS7 cells resulted in lower levels of 14-3-3 σ protein compared with cells transfected with 14-3-3 σ alone (Fig. 3d). Similar results were obtained in HEK293 cells (data not shown). To explore whether Efp overexpression can modulate 14-3-3 σ turnover, we performed pulse and chase experiments in transfected COS7 cells (Fig. 3e). In the vector-transfected cells, ³⁵S-labelled 14-3-3 σ immunoprecipitated by anti-haemagglutinin (HA) remains at more than 50% after 3 h, whereas the degradation rate of 14-3-3 σ was explicitly enhanced in Efp-overexpressed cells (12.2% 14-3-3 σ remaining after 3 h in Efp-overexpressed cells compared with 58.4% in control cells). The reduced levels of 14-3-3 σ in MCF7 cells caused by adenoviral Efp transduction were recovered by MG132, a potent inhibitor of proteasome function¹² (Fig. 3f). The finding that MG132 increases the amount of 14-3-3 σ binding to Efp was further confirmed by immunoprecipitation (Fig. 3g). These data suggest that Efp directly downregulates 14-3-3 σ levels through a proteasome-dependent mechanism.

We next investigated how Efp downregulates expression of 14-3-3 σ . One possible mechanism is that Efp functions as an E3 ubiquitin ligase similar to other RING finger proteins such as Rbx1 (refs

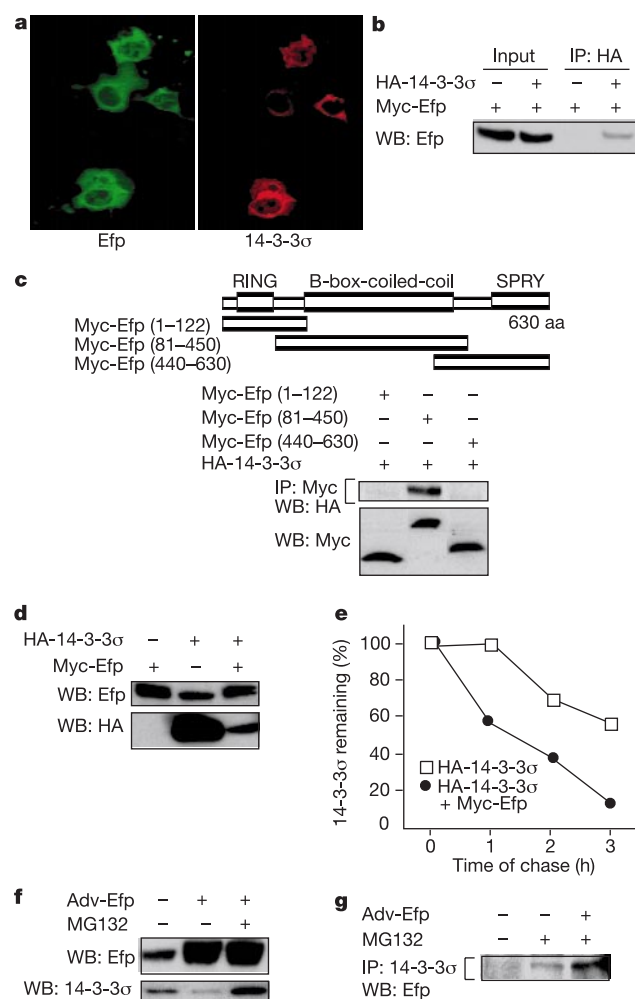


Figure 3 Efp conjugates to 14-3-3 σ and promotes 14-3-3 σ proteolysis in a proteasome-dependent manner. **a**, Immunoreactivities of Efp (fluorescein isothiocyanate (FITC), left panel) and 14-3-3 σ (rhodamin, right panel) expressed in COS7 cells were co-localized in the cytoplasm. **b**, Interaction of Efp with 14-3-3 σ . Lysates of COS7 cells transfected with indicated plasmids were analysed by immunoblotting using anti-Efp antibody (WB: Efp), combined without (input) or with (IP: HA) immunoprecipitation using anti-HA. **c**, Specific interaction of the B-box coiled-coil domain in Efp with 14-3-3 σ . Lysates of transfected COS7 cells were immunoprecipitated with anti-Myc and analysed by immunoblotting using anti-HA (middle panel: IP: Myc, WB: HA). The amounts of Myc-Efp mutants (WB: Myc) are shown in the bottom panel. **d**, Efp overexpression downregulates exogenous 14-3-3 σ . Lysates of transfected COS7 cells were analysed by anti-Efp (top panel) or anti-HA (bottom panel). **e**, 14-3-3 σ degradation is accelerated by Efp expression. COS7 cells transiently transfected with HA-14-3-3 σ and vector/Myc-Efp were pulse-labelled with [³⁵S]methionine for 30 min, and chased for the indicated time. **f**, Adenoviral Efp expression downregulates endogenous 14-3-3 σ in a proteasome-dependent manner. MCF7 cells transduced with β -galactosidase (Adv-Efp -) or Efp (Adv-Efp +) adenoviruses were incubated with either vehicle (MG132 -) or MG132 (10 μ M) for 8 h. Lysates were analysed by anti-Efp (top panel) or anti-14-3-3 σ (bottom panel). **g**, MG132 enhanced *in vivo* interaction between adenovirally expressing Efp and endogenous 14-3-3 σ . Lysates of MCF7 cells were immunoprecipitated with anti-14-3-3 σ and analysed by anti-Efp.

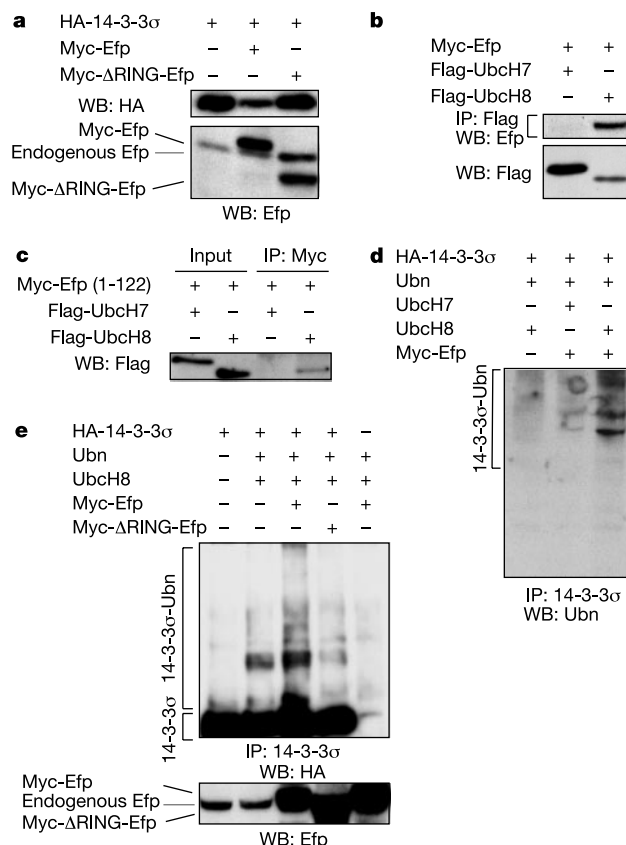


Figure 4 Efp is a RING-finger-dependent ubiquitin ligase that targets 14-3-3 σ . **a**, 14-3-3 σ downregulation is mediated through the RING finger in Efp. COS7 cells were co-transfected with HA-14-3-3 σ and the indicated Myc-Efp mutants. Top panel, 14-3-3 σ levels; bottom panel, expression of Efp with or without Myc-tagging. **b**, Association of Efp with Ubch8. COS7 cells were co-transfected with Myc-Efp and Flag-tagged Ubch7/Ubch8. **c**, The RING finger in Efp is required for interaction with Ubch8. COS7 cells were co-transfected with Myc-Efp (residues 1–122) and Flag-tagged Ubch7/Ubch8. **d**, Efp and Ubch8 promote 14-3-3 σ ubiquitination. Anti-ubiquitin was used to detect ubiquitinated forms in 14-3-3 σ immunoprecipitates obtained from transfected COS7 cells. **e**, The RING finger in Efp is required for 14-3-3 σ ubiquitination. Anti-HA was used to detect HA-14-3-3 σ and its ubiquitinated forms (Ubn) in 14-3-3 σ precipitates obtained from transfected COS7 cells (top panel). The bottom panel shows expression of Efp with or without Myc-tagging.

13, 14), BRCA1 (ref. 15), Cbl (ref. 16), Mdm2 (ref. 17), parkin (ref. 18) and MID1 (ref. 19). The ubiquitination-dependent proteolysis is related to cell cycle progression and carcinogenesis^{20–23}. We constructed a RING-finger-deleted Efp mutant that lacked the 31 N-terminal amino acids (Δ RING-Efp), and tested its function in COS7 cells. The Δ RING-Efp mutant failed to downregulate levels of 14-3-3 σ , indicating that the RING finger in Efp is essential for regulation of 14-3-3 σ (Fig. 4a). Furthermore, Efp preferentially bound to ubiquitin-conjugating enzyme UbcH8 rather than UbcH7 (Fig. 4b). This interaction is RING-finger-dependent because Efp (1–122) was immunoprecipitated with UbcH8 (Fig. 4c). We next examined whether Efp can ubiquitinate 14-3-3 σ in a UbcH8-dependent manner. COS7 cells were transfected with Efp, 14-3-3 σ , ubiquitin and UbcH8/UbcH7. Large amounts of ubiquitinated 14-3-3 σ were observed in cells transfected with UbcH8 but not with UbcH7 (Fig. 4d). Cells transfected with the Δ RING-Efp mutant failed to ubiquitinate 14-3-3 σ , confirming that ubiquitination requires the Efp RING finger (Fig. 4e). We conclude that Efp downregulates 14-3-3 σ through ubiquitin-mediated degradation in which Efp functions as an E3 that selectively requires UbcH8.

We investigated further the effects of Efp abrogation on cell growth using mouse embryonic fibroblasts (MEFs) from *Efp*^{+/+} and *Efp*^{-/-} mice as models. The growth of *Efp*^{+/+} MEFs was normal, whereas that of *Efp*^{-/-} MEFs was severely reduced (Fig. 5a). 14-3-3 σ levels were markedly elevated in *Efp*^{-/-} MEFs (Fig. 5b). The levels of Cdk2 and Cdc2 bound to p21^{Cip1} or 14-3-3 σ were

elevated in *Efp*^{-/-} MEFs (Fig. 5c), although the levels of Cdk2 and Cdc2 in both MEFs were not significantly different (Fig. 5c). This suggests that increased amounts of 14-3-3 σ and p21^{Cip1} in *Efp*^{-/-} MEFs tend to entrap mitotic factors, leading to reduced growth. We examined the effects of 14-3-3 σ antisense/sense oligonucleotides on MEF cell growth by collecting cells 24 or 48 h after initial incubation with oligonucleotides (Fig. 5d). The reduced cell growth in *Efp*^{-/-} MEFs was rescued by the antisense oligonucleotide in a dose-dependent manner, suggesting that 14-3-3 σ is the principal checkpoint in *Efp*^{-/-} MEFs. To explore whether the accumulation of 14-3-3 σ in *Efp*^{-/-} MEFs is due to stabilization of the protein rather than activation of transcription, we performed pulse and chase experiments (Fig. 5e). The half-life ($t_{1/2}$) of 14-3-3 σ was 4 h in *Efp*^{-/-} MEFs (95% confidence interval, 6.7–2.9 h), which was significantly longer than $t_{1/2}$ in *Efp*^{+/+} MEFs (2.3 h, 95% confidence interval, 2.8–1.8 h, $P < 0.001$). The accumulation of 14-3-3 σ in *Efp*^{-/-} MEFs was reversed dose-dependently by adenoviral Efp transduction (Fig. 5f). The reduced levels of 14-3-3 σ in *Efp*^{-/-} MEFs adenovirally expressing Efp was rescued by MG132, which is consistent with the results indicating that Efp degrades 14-3-3 σ through a ubiquitin-dependent pathway.

Our findings provide insight into the cell cycle machinery and growth of oestrogen-dependent tumours. In the classical scheme, DNA damage induces elevation of p53 levels and p53 promotes transcription of Cdk inhibitors, which recruit cyclin–Cdk complexes leading to cell cycle arrest and DNA repair^{8,24–27}. Here we

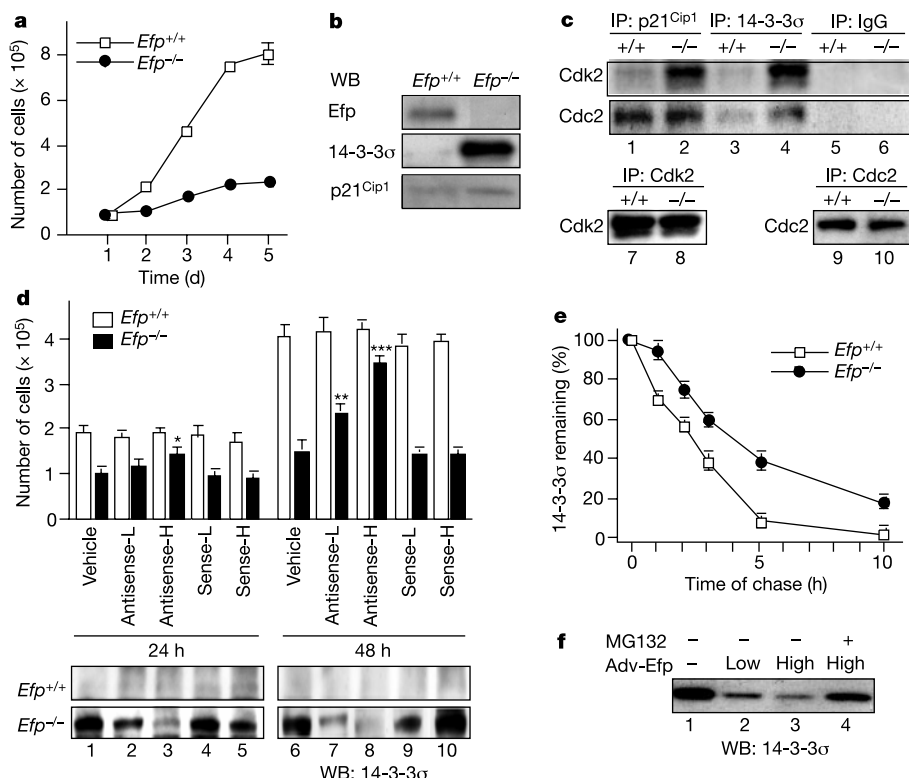


Figure 5 Loss of Efp function causes an accumulation of 14-3-3 σ and inhibits cell proliferation. **a**, The growth of *Efp*^{-/-} MEFs is reduced compared with *Efp*^{+/+} MEFs. **b**, Accumulation of 14-3-3 σ in *Efp*^{-/-} MEFs. **c**, High amounts of Cdk2 (top panel) or Cdc2 (middle panel) associated with p21^{Cip1} or 14-3-3 σ are detected in *Efp*^{-/-} MEFs. The amounts of precipitated Cdk2 and Cdc2 are shown in the bottom panel. **d**, Reduced growth in *Efp*^{-/-} MEFs was rescued by antisense 14-3-3 σ oligonucleotide. MEFs were incubated with antisense/sense oligonucleotides (antisense-L/sense-L, 100 nM; antisense-H/sense-H, 400 nM), or with vehicle. The graph shows the number of cells collected 24 or 48 h after initial oligonucleotide transfection. The bottom panel shows 14-3-3 σ levels. Lanes 1–5 and 6–10 show lysates from indicated cells collected 24 and

48 h after initial transfection, respectively. Asterisk, $P < 0.05$ compared with vehicle at 24 h; double asterisk, $P < 0.01$ compared with vehicle at 48 h; triple asterisk, $P < 0.001$ compared with vehicle at 48 h. **e**, 14-3-3 σ turnover is more stabilized in *Efp*^{-/-} MEFs than in *Efp*^{+/+} MEFs. Lysates were immunoprecipitated by anti-14-3-3 σ . **f**, 14-3-3 σ accumulation in *Efp*^{-/-} MEFs is reduced by adenoviral Efp expression in a proteasome-dependent manner. MEFs adenovirally expressing β -galactosidase (lane 1) or Efp at a low dose (40 MOI) (lane 2) or at a high dose (100 MOI) (lanes 3 and 4) were subjected to immunoblotting using anti-14-3-3 σ . Cells in lane 4 were treated with MG132 (10 μ M).

demonstrate a mechanism in which the function of 14-3-3 σ is modulated by Efp-mediated proteolysis. The degradation of 14-3-3 σ is subsequently followed by dissociation of the protein from cyclin-Cdk complexes, leading to cell cycle progression and tumour growth. Indeed, it has been shown recently that downregulation of 14-3-3 σ alone is sufficient to immortalize primary keratinocytes²⁸. The present data suggest that Efp may be an essential oncogenic factor in breast cancers and perhaps other hormone-related malignancies. Our findings have important clinical relevance because targeting the Efp-mediated breakdown of 14-3-3 σ could provide a new therapy to block breast tumour proliferation. □

Methods

MCF7 tumour growth in nude mice

MCF7 cells suspended in Matrigel (5×10^6 cells per 0.1 ml) were injected subcutaneously into female athymic mice (4-week-old BALB/c nu/nu). We calculated tumour volumes weekly by the tumour radii. When the tumour volume reached 300 mm³, mice were ovariectomized or treated with Efp sense/antisense oligonucleotides (phosphorothioate-modified). Oligonucleotide sequences are: antisense, 5'-AGGGGGCACAGCTCTGCCAT-3'; sense, 5'-ATGGCAGAGCTGTGCCCT-3'. Six weeks later, animals were killed and tumours were analysed by immunoblotting using anti-Efp antibody¹. In experiments for Efp-expressing MCF7 cells, cells were transfected with pEF²⁹ or pEF containing Efp. Two clones for each construct containing either pEF alone (vector-MCF7 number 1 and 3) or pEF-Efp (Efp-MCF7 number 2 and 4) were selected by G418. Each cell suspension (5×10^6 cells per 0.1 ml) was injected into five ovariectomized nude mice respectively. In other experiments, five intact female athymic mice injected with MCF7 cells were ovariectomized at a tumour volume of 180 mm³.

Cell cycle analysis

Transfected MCF7 cells were incubated with phenol red free DMEM containing 0.2% charcoal-stripped fetal bovine serum (FBS) 24 h before analysis. Trypsinized cells were resuspended in hypotonic propidium iodide solution ($50 \mu\text{g ml}^{-1}$) containing 0.1% sodium citrate and 0.1% Triton X-100, and analysed on a FACScan flow cytometer (Becton Dickinson).

Yeast two-hybrid screen

Saccharomyces cerevisiae strain EGY48 was transformed with pEG202-NLS-human Efp, and further transformed with a mouse embryo cDNA library (OriGene). Selected positive clones according to the manufacturer's instructions were sequenced.

Immunoprecipitation assays

In experiments with COS7 cells, cells were lysed in Nonidet P-40 lysis buffer³⁰ 48 h after plasmid transfection. Lysates were immunoprecipitated with anti-Myc (Santa Cruz), anti-Flag (Sigma), or anti-14-3-3 σ (Santa Cruz) antibodies. In experiments with MEFs, lysates were immunoprecipitated with anti-p21^{Cip1} (Santa Cruz), anti-14-3-3 σ , or goat immunoglobulin- γ (IgG), and immunoblotted by anti-Cdk2 (Santa Cruz) and anti-Cdc2 (Santa Cruz).

Pulse and chase experiments

MEFs were pulse-labelled with [³⁵S]methionine ($200 \mu\text{Ci ml}^{-1}$) in methionine-free medium containing 10% dialysed FBS for 2 h, washed twice with pre-warmed PBS, and chased by culturing in DMEM containing 10% FBS for 10 h. Lysates from pulse-labelled cells were immunoprecipitated with anti-14-3-3 σ , resolved by SDS-polyacrylamide gel electrophoresis, and analysed with a PhosphorImager (Molecular Dynamics). In experiments with COS7 cells, cells were pulse-labelled for 30 min and chased for 3 h. Lysates were immunoprecipitated with anti-HA (Santa Cruz).

Generation of recombinant adenoviruses

Recombinant adenoviruses were generated by Adenovirus Expression Vector Kit (TaKaRa) and used at a multiplicity of infection (MOI) of 40–100 according to the manufacturer's protocol.

In vivo ubiquitination assays

COS7 cells were transfected with plasmids including Myc-Efp, HA-14-3-3 σ , ubiquitin, and Flag-tagged UbCH8 or UbCH7. Forty hours later, cells were incubated with MG132 ($10 \mu\text{M}$, Peptide Institute) for 8 h, then immunoprecipitated with anti-14-3-3 σ . Precipitates were analysed by anti-ubiquitin (Santa Cruz) or anti-HA.

Preparation of Efp^{+/+} and Efp^{-/-} MEFs

We generated Efp^{-/-} mice as described⁴. MEFs were obtained from wild-type or Efp^{-/-} 13.5-day-old embryos and used at third to sixth generations. For growth rate analysis, cells were synchronized for 72 h in DMEM containing 0.1% FBS, then trypsinized and plated at a density of 1×10^5 cells per 35-mm dish in DMEM containing 10% FBS. In experiments for antisense oligonucleotide transfection, MEFs maintained for 24 h in DMEM containing 10% FBS were incubated with oligofectamine (Invitrogen) and either antisense/sense 14-3-3 σ oligonucleotides (phosphorothioate-modified) or vehicle alone for 4 h according to the manufacturer's protocol. At the end of incubation, 10% FBS was supplemented and maintained for the next 24 or 48 h. Antisense, 5'-ATCAGACTGGC

TCTCTCCAT-3'; sense, 5'-ATGGAGAGAGCCAGTCTGAT-3'. In other experiments, cells were seeded and incubated with medium containing 10% FBS for 48 h, then lysed in Nonidet P-40 lysis buffer³⁰.

Received 6 February; accepted 19 April 2002; doi:10.1038/nature00826.

- Inoue, S. *et al.* Genomic binding-site cloning reveals an estrogen-responsive gene that encodes a RING finger protein. *Proc. Natl Acad. Sci. USA* **90**, 11117–11121 (1993).
- Orimo, A., Inoue, S., Ikeda, K., Noji, S. & Muramatsu, M. Molecular cloning, structure, and expression of mouse estrogen-responsive finger protein Efp. Co-localization with estrogen receptor mRNA in target organs. *J. Biol. Chem.* **270**, 24406–24413 (1995).
- Muramatsu, M. & Inoue, S. Estrogen receptors: how do they control reproductive and non-reproductive functions. *Biochem. Biophys. Res. Commun.* **270**, 1–10 (2000).
- Freemont, P. S. RING for destruction? *Curr. Biol.* **10**, R84–R87 (2000).
- Ikeda, K., Orimo, A., Higashi, Y., Muramatsu, M. & Inoue, S. Efp as a primary estrogen-responsive gene in human breast cancer. *FEBS Lett.* **472**, 9–13 (2000).
- Orimo, A. *et al.* Underdeveloped uterus and reduced estrogen responsiveness in mice with disruption of the estrogen-responsive finger protein gene, which is a direct target of estrogen receptor. *Proc. Natl Acad. Sci. USA* **96**, 12027–12032 (1999).
- Hermeking, H. *et al.* 14-3-3 σ is a p53-regulated inhibitor of G2/M progression. *Mol. Cell* **1**, 3–11 (1997).
- Bunz, F. *et al.* Requirement for p53 and p21 to sustain G2 arrest after DNA damage. *Science* **282**, 1497–1501 (1998).
- Prasad, G. L., Valverius, E. M., McDuffie, E. & Cooper, H. L. Complementary DNA cloning of a novel epithelial cell marker protein, HME1, that may be down-regulated in neoplastic mammary cells. *Cell Growth Differ.* **3**, 507–513 (1992).
- Jensen, K., Shields, C. & Freemont, P. S. PML protein isoforms and the RBCC/TRIM motif. *Oncogene* **20**, 7223–7233 (2001).
- Le Douarin, B. *et al.* The N-terminal part of TIF1, a putative mediator of the ligand-dependent activation function (AF-2) of nuclear receptors, is fused to B-raf in the oncogenic protein T18. *EMBO J.* **14**, 2020–2033 (1995).
- Traenckner, E. B., Wilk, S. & Baeuerle, P. A. A proteasome inhibitor prevents activation of NF- κ B and stabilizes a newly phosphorylated form of I κ B- α that is still bound to NF- κ B. *EMBO J.* **13**, 5433–5441 (1994).
- Skowyra, D. *et al.* Reconstitution of G1 cyclin ubiquitination with complexes containing SCFGrr1 and Rbx1. *Science* **284**, 662–665 (1999).
- Kamura, T. *et al.* Rbx1, a component of the VHL tumour suppressor complex and SCF ubiquitin ligase. *Science* **284**, 657–661 (1999).
- Ruffner, H., Joazeiro, C. A., Hemmati, D., Hunter, T. & Verma, I. M. Cancer-predisposing mutations within the RING domain of BRCA1: loss of ubiquitin protein ligase activity and protection from radiation hypersensitivity. *Proc. Natl Acad. Sci. USA* **98**, 5134–5139 (2001).
- Joazeiro, C. A. *et al.* The tyrosine kinase negative regulator c-Cbl as a RING-type, E2-dependent ubiquitin-protein ligase. *Science* **286**, 309–312 (1999).
- Honda, R. & Yasuda, H. Association of p19 (ARF) with Mdm2 inhibits ubiquitin ligase activity of Mdm2 for tumour suppressor p53. *EMBO J.* **18**, 22–27 (1999).
- Shimura, H. *et al.* Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. *Nature Genet.* **25**, 302–305 (2000).
- Trockenbacher, A. *et al.* MID1, mutated in Opitz syndrome, encodes an ubiquitin ligase that targets phosphatase 2A for degradation. *Nature Genet.* **29**, 287–294 (2001).
- Tyers, M. & Willems, A. R. One ring to rule a superfamily of E3 ubiquitin ligases. *Science* **284**, 601–604 (1999).
- Joazeiro, C. A. P. & Weissman, A. M. RING finger proteins: mediators of ubiquitin ligase activity. *Cell* **102**, 549–552 (2000).
- Pagano, M. *et al.* Role of the ubiquitin-proteasome pathway in regulating abundance of the cyclin-dependent kinase inhibitor p27. *Science* **269**, 682–685 (1995).
- Pickart, C. M. Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* **70**, 503–533 (2001).
- Chan, T. A., Hermeking, H., Lengauer, C., Kinzler, K. W. & Vogelstein, B. 14-3-3 σ is required to prevent mitotic catastrophe after DNA damage. *Nature* **401**, 616–620 (1999).
- Laronga, C., Yang, H. Y., Neal, C. & Lee, M. H. Association of the cyclin-dependent kinases and 14-3-3 σ negatively regulates cell cycle progression. *J. Biol. Chem.* **275**, 23106–23112 (2000).
- Kastan, M. B. *et al.* A mammalian cell cycle checkpoint pathway utilizing p53 and GADD45 is defective in ataxia-telangiectasia. *Cell* **71**, 587–597 (1992).
- Cox, L. S. & Lane, D. P. Tumour suppressors, kinases and clamps: how p53 regulates the cell cycle in response to DNA damage. *Bioessays* **17**, 501–508 (1995).
- Dellambra, E. *et al.* Downregulation of 14-3-3 σ prevents clonal evolution and leads to immortalization of primary human keratinocytes. *J. Cell Biol.* **149**, 1117–1129 (2000).
- Kim, D. W., Uetsuki, T., Kaziro, Y., Yamaguchi, N. & Sugano, S. Use of the human elongation factor 1 α promoter as a versatile and efficient expression system. *Gene* **91**, 217–223 (1990).
- Urano, T. *et al.* p57(Kip2) is degraded through the proteasome in osteoblasts stimulated to proliferate by transforming growth factor- β 1. *J. Biol. Chem.* **274**, 12197–12200 (1999).

Acknowledgements

We thank H. Yashiroda, A. Takada, K. Okamoto, M. Kabayashi, S. Inada and T. Hishinuma for technical assistance, and B. Blumberg and K. Horie for critical reading and comments on the manuscript. This work was supported in part by grants-in-aid from the Ministry of Health, Labor and Welfare, from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.I. (e-mail: INOUE-GER@h.u-tokyo.ac.jp).

Crystal structure of parallel quadruplexes from human telomeric DNA

Gary N. Parkinson, Michael P. H. Lee & Stephen Neidle

The Cancer Research UK Biomolecular Structure Unit, Chester Beatty Laboratories, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK

Telomeric ends of chromosomes, which comprise noncoding repeat sequences of guanine-rich DNA, are fundamental in protecting the cell from recombination and degradation¹. Disruption of telomere maintenance leads to eventual cell death, which can be exploited for therapeutic intervention in cancer. Telomeric DNA sequences can form four-stranded (quadruplex) structures^{2–4}, which may be involved in the structure of telomere ends⁵. Here we describe the crystal structure of a quadruplex formed from four consecutive human telomeric DNA repeats and grown at a K^+ concentration that approximates its intracellular concentration. K^+ ions are observed in the structure. The folding and appearance of the DNA in this intramolecular quadruplex is fundamentally different from the published Na^+ -containing quadruplex structures^{2,4,6}. All four DNA strands are parallel, with the three linking trinucleotide loops positioned on the exterior of the quadruplex core, in a propeller-like arrangement. The adenine in each TTA linking trinucleotide loop is swung back so that it intercalates between the two thymines. This DNA structure suggests a straightforward path for telomere folding and unfolding, as well as ways in which it can recognize telomere-associated proteins.

Telomeric DNA of vertebrates consists of tandem repeats of the sequence d(TTAGGG), and in human somatic cells is typically 5–8 kilobases (kb) long, with a single-stranded 3' overhang of 100–200 bases⁷. These G-rich ends can fold up into four-stranded G-quadruplex structures⁸. Biophysical and structural methods have shown that the central units of G-quadruplexes are hydrogen-bonded arrays of guanine bases (G-quartets), with several G-quartets held together by π - π stacking interactions. Intermolecular quadruplexes are formed by two (dimeric) or four (unimolecular) separate strands associating together⁹. Intramolecular quadruplexes are formed by the folding of either two or four consecutive repeats, such as the four repeats of human telomeric DNA in d[AGGG(TTAGGG)₃]. The human single-stranded telomere end, which can potentially fold into a number of four-repeat quadruplexes, is far longer and more stable than in many other vertebrates. The potential role of quadruplexes *in vivo* has been highlighted with the recent development of therapeutic strategies designed to stabilize telomeric ends as G-quadruplex structures using specific small molecules, which can destabilize telomere maintenance in tumour cells^{10–13}. The characterization of a human nuclease with G-quadruplex specificity suggests that these structures may be involved as intermediates in recombination at G-rich sequences¹⁴. This is consistent with the finding that the same small-molecule ligands also inhibit the unwinding of G-quadruplexes by helicases that are involved in recombination¹⁵.

We report here the crystal structures of both intra- and intermolecular K^+ -containing quadruplexes formed from 12-nucleotide and 22-nucleotide repeat sequences ('12- and 22-mers') from human telomeres. These reveal a remarkable and completely unexpected folding topology and overall structural type. Both quadruplexes were expected to consist of four strands alternating between parallel and antiparallel orientations, with the TTA loops connecting G-quartets. The structures that have been found for

quadruplexes from *Oxytricha nova*^{2,4,16}, and in a nuclear magnetic resonance (NMR) analysis of a Na^+ -containing human four-repeat 22-mer sequence⁶, have diagonal and lateral loops at quadruplex ends (Fig. 1c). Instead we find that both the dimeric (12-mer) intermolecular and the 22-mer intramolecular quadruplex each have all four strands in a parallel arrangement. The resulting topology, common to both structures, has the TTA loops required to be extended out from the sides of the stacks of three consecutive G-quartets (Fig. 1a, b). So under crystallizing conditions containing K^+ ions, which is analogous to the intracellular ionic environment, human telomeres fold in a manner that is fundamentally distinct from those of lower organisms. Earlier circular dichroism studies on intramolecular human telomeric quadruplex sequences suggested¹⁷ that the antiparallel fold is solely a consequence of Na^+ ions, whereas the more strongly bound K^+ ions induce a transition to a parallel arrangement, although this was not defined in structural terms.

The sequences studied, d(TAGGGTTAGGGT) and d[AGGG(TTAGGG)₃], both form quadruplexes that have approximately four-fold noncrystallographic symmetry for the central core of three G-quartets. The backbone of each 12-mer has one TTA moiety, which forms a loop that projects outwards from the semicircle of the backbone (Fig. 2a, b). The loop connects the top of one strand with the bottom of the other, ensuring their parallelism and the integrity of the G-quartets. Thus the resulting quadruplex has two such loops, opposite to one another. All the guanine glycosidic bonds have an *anti* conformation, with C2'-*endo* sugar puckers. The 12 guanines from the two strands are arranged into three stacked G-quartets that generate a structure (Fig. 2a–c) resembling a flattened disc, 41 Å wide and 6.3 Å tall. The two TTA loops, one from each strand, protrude outwards like the blades of a propeller. The G-quartets in the 22-mer intramolecular quadruplex have an identical arrangement, and again the strand polarities are all

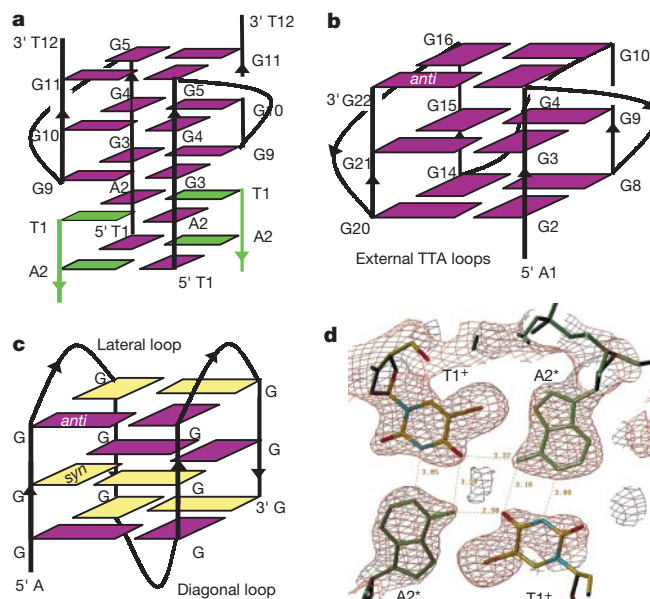


Figure 1 Schematic diagram of human telomeric quadruplex folding topologies. **a, b**, Intramolecular and dimeric intermolecular G-quadruplexes, all with *anti* glycosidic torsion angles and extended external loops abutting the sides of the G-quadruplex crystallized with K^+ ions. **c**, Fold determined from the NMR intramolecular G-quadruplex solution structure⁶ in Na^+ ions only, with lateral and diagonal loops, of sequence d(AG₃(T₂AG₃)). Yellow boxes denote *syn* guanines. **d**, Final ($2F_o - F_c$) electron density around the T⁺AT⁺A quartet. The σ_A -weighted map, calculated using 10–2.4 Å data, was contoured at 1.2 σ .

parallel. There are three TTA groups in this sequence, and all three form these TTA loops. The morphology of this quadruplex is thus closely similar to that of the 12-mer dimer, except that there are now three protruding loops, enhancing the propeller-like appearance of the quadruplex (Fig. 3b, d, e).

The G-quartets in both quadruplexes have the characteristic square planar arrangement found in simple quadruplexes, with base pairing through their Watson–Crick and Hoogsteen edges. The local G-quartet rise is 3.13 Å, with an average 30° twist between successive ones. Monovalent K^+ ions are positioned between the stacked G-quartets, 2.7 Å from each of the eight O6 carbonyl groups, in a bipyramidal antiprismatic arrangement. The backbone torsion angles in both human quadruplex structures are all in standard DNA ranges except for dihedral angles ζ (65°), α (74°) and γ (186°), located at a stacked G-to-TTA loop interface. The loops are involved in hydrogen bonding and stacking interactions with symmetry-related molecules in the crystal lattices. Remarkably, these have little effect on conformation because all the loops in both structures are very similar, with the adenine in each TTA sequence

swung back so that it intercalates between the two thymines. This results in the second thymine (thymine2) in each loop adopting a C3'-endo sugar pucker, whereas the other thymine and the adenine both adopt C2'-endo puckers. Each thymine2 is positioned at the tip of the loop, such that it can interact in a variety of ways with other molecules. Its interactions observed in the crystal lattice include stacking with a guanine and thymine...thymine O2...N3 base pairing. This suggests that these loops are used by quadruplexes to make intermolecular interactions, for example with telomeric proteins.

Because the loops extend laterally up to 10 Å from the core G-quartets, the TTA loops in both structures generate three very different categories of surface (Fig. 4a): a polarized surface, a hydrophobic aromatic planar surface and a TAT loop hydrogen-bonding interface, as described above. The 12-mer quadruplex has one planar surface involved with the stacking of a pair of adenines, and above them a pair of thymines from the 5' ends of each strand. These are related by a crystallographic two-fold axis, which results in them becoming two stacked, symmetry-equivalent TATA

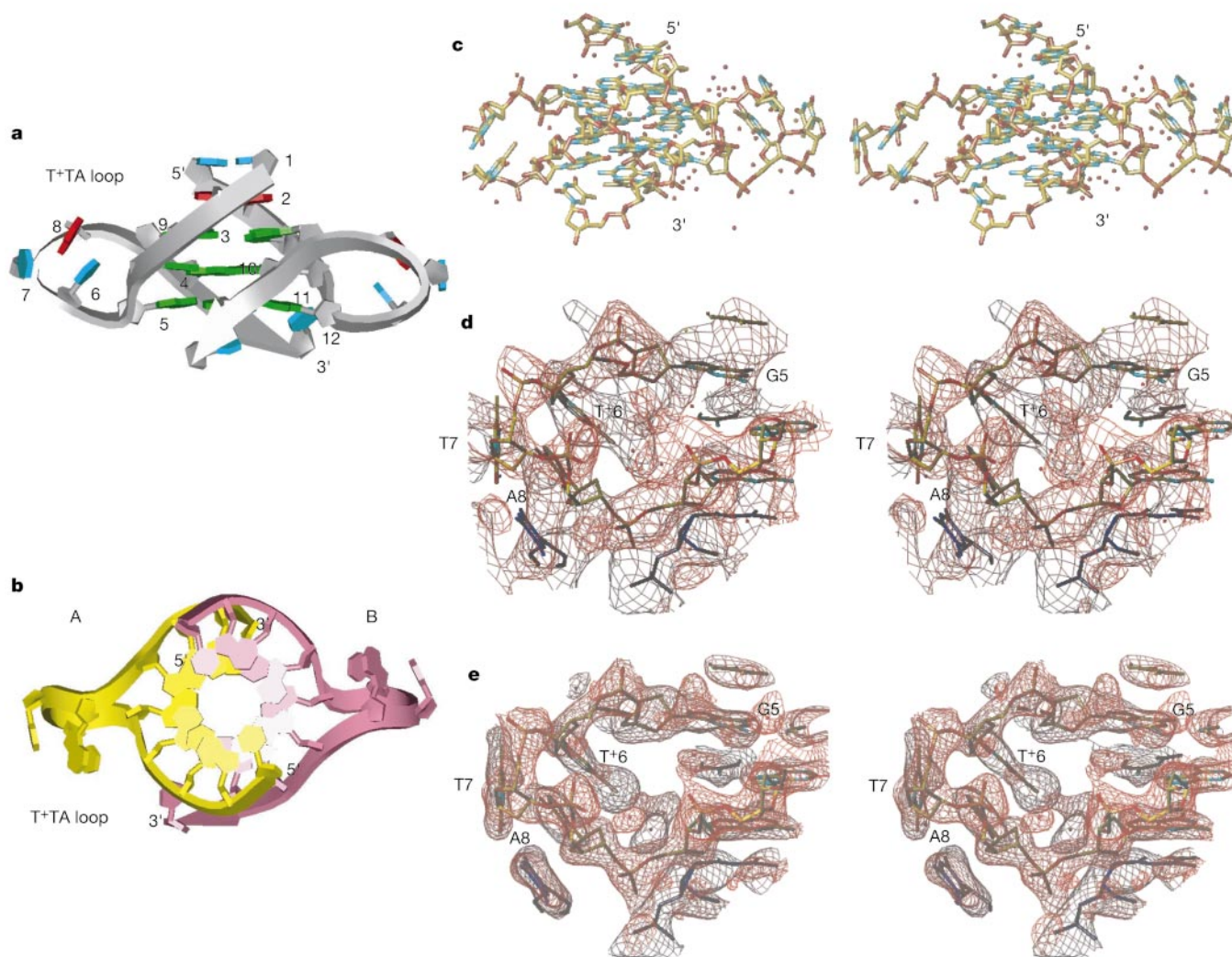


Figure 2 The overall folding topology of the two 12-mers constituting the content of the asymmetric unit in the dimeric quadruplex. **a**, Side view of the quadruplex with the phosphate sugar backbone drawn as a ribbon showing 5'–3' directionality. The guanines are green, thymines and modified uracils are blue, and adenines are red. **b**, View from the 5' end of the quadruplex. Strands A and B are yellow and magenta, respectively. **c**, Stereo stick representation coloured by atom type. Drawn with the

program TURBO²⁹. **d**, The initial electron-density map calculated with MAD solvent-flattened phases to 2.9 Å resolution, contoured at a 1.6 σ level and centred around the extended TTA loop region abutting the sides of the G-quadruplex. The final model is overlaid for clarity. **e**, The same region showing the final α_A -weighted $2F_O - F_C$ electron density map contoured at 1.4 σ (using data at 10–2.4 Å resolution, drawn with the program TURBO²⁹).

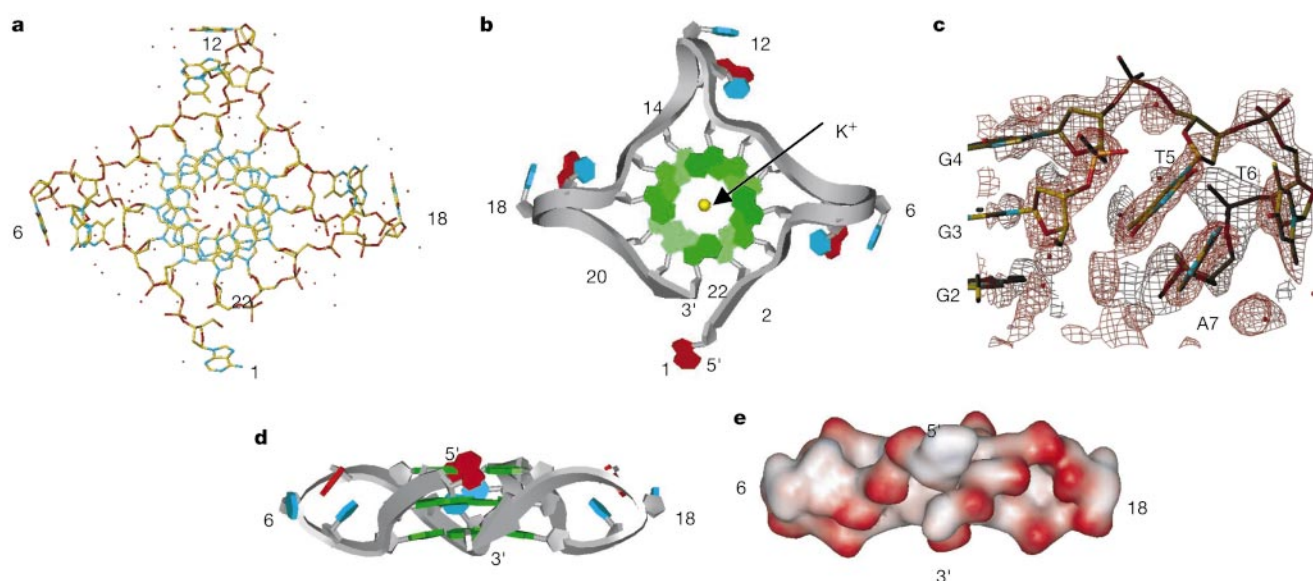


Figure 3 Overall folding topology of the 22-mer intramolecular G-quadruplex. **a**, Stick representation coloured by atom type and viewed on the 5' face. The central potassium counter ion is coordinated in a bipyramidal antiprismatic arrangement by the electronegative carbonyl groups of guanine O6. Drawn with the program TURBO²⁹. **b**, View from the 3' end of the quadruplex looking down the helical axis with the phosphate sugar backbone drawn as a grey ribbon showing 5'-to-3' directionality.

Guanines are green, thymines blue, and adenines red. **c**, A representative part of the structure around the extended TTA loop region abutting the sides of the G-quadruplex. Overlaid is a α_A -weighted map using data at 10–2.1 Å resolution contoured at 1.8 σ . **d**, Side view of the quadruplex highlighting its disc-like shape and positioning of the 3' and 5' strand ends. **e**, Space-filling van der Waals contoured representation, coloured by charge, with red surfaces representing regions of negative charge.

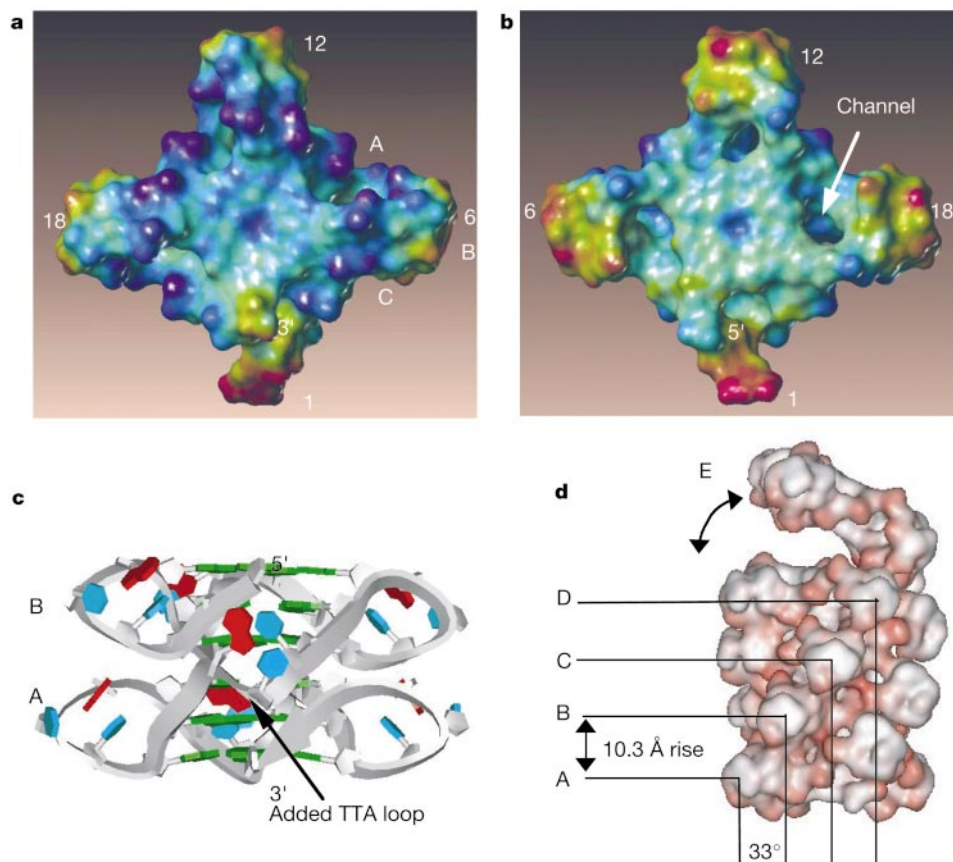


Figure 4 Visualizations of the 22-mer and potential 22-mer repeats. **a**, Space-filling van der Waals contoured visualization. Solvent atoms are removed, and view is onto the 3' G tetrad plane. **b**, View onto the 5' G-tetrad plane highlighting the differences between the 5' and 3' surfaces and the channels between the TTA loops and the G tetrads, coloured by charge. **c**, Model showing two telomeric human G-quadruplex repeats stacked 3' to 5'.

The upper stack has been rotated 33° relative to the lower quadruplex, with an unmodified TTA loop modelled between the two to link them. **d**, A model for higher-order telomeric DNA structure at the end of a human chromosome. Four quadruplex repeats have been stacked using the same building method employed in **c**. A fifth quadruplex repeat is shown linked and folding onto the stack.

quartets, with the adenines and thymines from a second, two-fold related G-quadruplex. The TATA (Fig. 1d) quartet also has its bases hydrogen bonded through their Watson–Crick and Hoogsteen edges, with conventional A•T Watson–Crick base pairing together with hydrogen bonds between N6(adenine) and O4(thymine). The counter-ion between the terminal G-quartet and the TATA quartet of the dimeric intermolecular G-quadruplex is Na^+ , which does not display the same coordination geometry as the K^+ ions in these structures. End-to-end stacking through the planar TATA quartets results in a stacked 5'-to-5' dimeric bimolecular quadruplex, with the TTA loops of the second symmetry-related G-quadruplex rotated by about 45° around the helical axis away from the TTA loops below. Packing of the 22-mer intramolecular quadruplex in the lattice also occurs through a 5'-to-5' arrangement, by directly involving the quadruplex G-quartets; this hydrophobic 5' face is different from the more hydrophilic 3' face (Fig. 4a, b).

All four grooves between the phosphate backbones contain several water molecules that hydrogen bond to sugar O3' and phosphate O1P atoms as well as to guanine N2 atoms. Only slight variations in groove width (9.0–10.3 Å) are observed between the four, apparently dependent on the presence or absence of a TTA linkage. The presence of the loops in the grooves gives them a distinct character. Rather than being continuous from one end of the G-quartet stack to the other, these grooves are finite, V-shaped, and have walls that do not simply comprise phosphate–sugar backbones. Solvent molecules form clustered networks that reflect this increased groove complexity relative to the simpler spines of hydration observed in the *Oxytricha* quadruplex crystal structure⁴.

Many characteristics of ligands that stabilize G-quadruplexes can now be rationalized in the light of this new structure for human quadruplexes. These ligands, typified by substituted acridines^{13,18,19} tetra-*N*-pyridyl porphyrins²⁰, triazine and ethidium derivatives^{21,22}, share common features of large planar surface areas together with cationic side chains. NMR and molecular modelling studies have suggested that these ligands interact at the ends of the G-quartet stacks^{23,24}. This interaction can readily occur through π -stacking, unhindered by the need to change lateral and diagonal loop conformation to do so, which would be the case with antiparallel or mixed parallel and antiparallel quadruplexes. There are four equivalent phosphate grooves for the binding of ligand substituents, together with the extra cavities adjacent to the loops. Modelling studies suggest that the potent trisubstituted acridine ligand earlier designed and synthesized by us¹³ is consistent with these features. At the same time, the complexity of the loop organization may suggest new types of ligand, whose features will be very different from existing ones, and thus may confer greater specificity. The simpler fold of our parallel structures, compared with antiparallel structural models, suggests an obvious pathway for readily folding and unfolding G-quadruplex structures. Thus the rapid folding kinetics noted for the human two-repeat dimer when binding to a ligand²⁵ is readily accommodated by the present structure, compared with an antiparallel hairpin one. The hypothesis that quadruplex types of structures are intermediates in recombination would require such a facile folding and unfolding. Previously observed topologies, as seen in the *Oxytricha* and Na^+ -containing four-repeat human quadruplexes, present a topological problem when attempting to model extended quadruplex telomeric sequences. Knots would quickly form when folding or unfolding these longer sequences. The structures revealed here do not present any such topological difficulties when extended to longer telomeric sequences. Indeed, the oligomerization of individual quadruplexes can be simply performed by inserting an additional fourth TTA loop to each 22-mer structure (Fig. 4c, d). A 200-base pair (bp) telomeric DNA sequence, if folded into a stack of quadruplexes, would form a cylindrical quasi-superhelix roughly 60 Å long (compared with a 680 Å-long B-DNA helix). The loops form a regular array on the

exterior of the superhelix, suitable for interaction with telomeric proteins such as TRF1 or with the nuclear envelope, or for inhibiting telomerase extension, especially when stabilized by ligand binding. □

Methods

DNA crystallization

DNA sequences d(T⁺AGGGT⁺TAGGGT)—where T⁺ is 5-bromouracil—and d(AGGGT⁺AGGGT⁺TAGGGT⁺TAGGGT) were purchased from the Oswel DNA Service (Southampton University). Before use, DNAs were heated to 358 K for 5 min and annealed to room temperature overnight in 50 mM potassium cacodylate buffer at pH 6.5. High-resolution DNA crystals for the sequence d(T⁺AGGGT⁺TAGGGT) were grown by vapour diffusion from hanging drops at 285 K using a 50% ammonium sulphate gradient. The initial drop conditions were 500 mM $(\text{NH}_4)_2\text{SO}_4$, 50 mM NaCl, 50 mM KCl, 50 mM Li_2SO_4 , 0.5 mM DNA and 50 mM potassium cacodylate buffered at pH 6.5. The crystals grew over several weeks as large hexagonal rods of dimensions $0.2 \times 0.2 \times 0.4$ mm. The DNA sequence d(AGGGT⁺AGGGT⁺TAGGGT⁺TAGGGT) was purified by anion-exchange chromatography (HQ/M Poros) and buffer exchanged into 50 mM potassium cacodylate at pH 6.5 with 30 mM KCl. High-resolution DNA crystals of the 22-mer were grown by vapour diffusion from hanging drops at 285 K using a 75% gradient. The initial drop conditions were 300 mM KI, 15% v/v polyethylene glycol (PEG) 400, 1.7 mM DNA, 2 mM BRACO19 trisubstituted acridine derivative, and 50 mM potassium cacodylate buffered at pH 6.5. The crystals grew over several weeks as large hexagonal rods of dimensions $0.1 \times 0.1 \times 0.4$ mm.

Crystallography of the 12-mer strand

The structure of the sequence d(T⁺AGGGT⁺TAGGGT) was determined using experimental phases obtained from a three-wavelength anomalous diffraction (MAD) experiment using the 5-bromouracil nucleotides as anomalous scatterers. Three data sets at 2.6–2.9 Å and at three different wavelengths were collected using a single flash-frozen crystal at 100 K. MAD data were collected on beamline ID14.4 at the European Synchrotron Radiation Facility (<http://www.esrf.fr>). A 2.4 Å data set collected in-house with a RAXIS IV image plate detector was used for the high-resolution refinement. Data were processed and scaled with the DENZO and SCALEPACK programs²⁶. Two of the four bromine atoms were identified from anomalous Patterson maps and used in the structure determination. Subsequent difference Fourier maps revealed one additional Br site in the T⁺TA loop of strand A. Positions and occupancies were refined with the CNS program²⁷ and the experimental phases were solvent corrected and extended to 2.7 Å. A partial model was built into the experimental maps. Interpretable Fourier maps were obtained using 67% of the model, allowing the remaining residues in the asymmetric unit to be unambiguously located. Refinement was initially performed with CNS using standard protocols, then after further model building, was continued with the Shelx-97 program²⁸, and extended to 2.4 Å resolution. Only the T⁺TA loop of strand A was visible in the initial solvent flattened maps and final α_A -weighted $2F_O - F_C$ and $F_O - F_C$ density maps (Fig. 2d, e). The nucleotides forming the second T⁺TA loop, for strand B, modelled on loop A, were included only during the latter part of the refinement, when electron density was observed and given 70% occupancy. The model was refined with all solvent molecules included, using data at 10–2.4 Å, to R and R_{free} values of 18.5% and 26.5% respectively.

Crystallography of the 22-mer quadruplex

Data for the 22-mer human G-quadruplex were collected using a single flash-frozen crystal at 100 K, again on beamline ID14.2. The crystal was first cryoprotected in 500 mM KI, 15% v/v PEG 400, 50 mM potassium cacodylate buffered at pH 6.5, and 17% glycerol. Data were processed and scaled as before. The structure was determined by molecular replacement using CNS with the G-quartet core of the dimeric G-quadruplex structure (73% of the scattering mass) as a starting point. Rigid-body refinement was followed by minimization with CNS and iterative cycles of manual fitting. Subsequent difference Fourier maps revealed all three TTA loops not included in the initial model. Further model building and refinement with Shelx-97 was performed to 2.1 Å resolution. All three of the TTA loops are clearly visible in the α_A -weighted $2F_O - F_C$ density maps (Fig. 3c). The final nucleotide A1 hydrogen bonds to a symmetry-related TTA loop that was included during the latter part of the refinement. The model was further refined with solvent molecules included, using data at 10–2.1 Å, to R and R_{free} values of 22.6% and 26.5% respectively. No residual electron density was observed that could account for the trisubstituted acridine derivative. Refinement statistics are given in the Supplementary Information.

Received 14 December 2001; accepted 10 May 2002; doi:10.1038/nature755.

Published online 26 May 2002.

- Hackett, J. A., Feldser, D. M. & Greider, C. W. Telomere dysfunction increases mutation rate and genomic instability. *Cell* **106**, 275–286 (2001).
- Smith, F. W. & Feigon, J. Quadruplex structure of oxytricha telomeric DNA oligonucleotides. *Nature* **356**, 164–168 (1992).
- Wang, Y. & Patel, D. J. Guanine residues in d(T₂AG₃) and d(T₂G₄) form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution. *Biochemistry* **31**, 8112–8119 (1992).
- Horvath, M. P. & Schultz, S. C. DNA G-quartets in a 1.86 Å resolution structure of an *Oxytricha nova* telomeric protein–DNA complex. *J. Mol. Biol.* **310**, 367–377 (2001).
- Schaffitzel, C. et al. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylylonchia lemnae* macronuclei. *Proc. Natl Acad. Sci. USA* **98**, 8572–8577 (2001).

6. Wang, Y. & Patel, D. J. Solution structure of the human telomeric repeat d[AG₃(T₂AG₃)₃] G-tetraplex. *Structure* **1**, 263–282 (1993).
7. Wright, W. E. *et al.* Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes Dev.* **11**, 2801–2809 (1997).
8. Simonsson, T. G-quadruplex DNA structures—variations on a theme. *Biol. Chem.* **382**, 621–628 (2001).
9. Phillips, K., Dauter, Z., Murchie, A. I., Lilley, D. M. & Luisi, B. The crystal structure of a parallel-stranded guanine tetraplex at 0.95 Å resolution. *J. Mol. Biol.* **273**, 171–182 (1997).
10. Mergny, J.-L. & Hélène, C. G-quadruplex DNA: a target for drug design. *Nature Genet.* **4**, 1366–1367 (1998).
11. Bearss, D. J., Hurley, L. H. & Von Hoff, D. D. Telomere maintenance mechanisms as a target for drug development. *Oncogene* **19**, 6632–6641 (2000).
12. Gowan, S. M. *et al.* A G-Quadruplex-interactive potent small-molecule inhibitor of telomerase exhibiting in vitro and in vivo antitumor activity. *Mol. Pharmacol.* **61**, 1154–1162 (2002).
13. Read, M. A. *et al.* Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors. *Proc. Natl Acad. Sci. USA* **98**, 4844–4849 (2001).
14. Sun, H., Yabuki, A. & Maizels, N. A human nuclease specific for G4 DNA. *Proc. Natl Acad. Sci. USA* **98**, 12444–12449 (2001).
15. Li, J.-L. *et al.* Inhibition of the Bloom's and Werner's syndrome helicases by G-quadruplex interacting ligands. *Biochemistry* **40**, 15194–15202 (2001).
16. Smith, F. W., Schultze, P. & Feigon, J. Solution structures of unimolecular quadruplexes formed by oligonucleotides containing *Oxytricha* telomere repeats. *Structure* **3**, 997–1008 (1995).
17. Balagurumoorthy, P. B. & Brahmachari, S. K. Structure and stability of human telomeric sequence. *J. Biol. Chem.* **269**, 21858–21869 (1994).
18. Read, M. A. *et al.* Molecular modeling studies on G-quadruplex complexes of telomerase inhibitors: Structure-activity relationships. *J. Med. Chem.* **42**, 4538–4546 (1999).
19. Harrison, R. J., Gowan, S. M., Kelland, L. R. & Neidle, S. Human telomerase inhibition by substituted acridine derivatives. *Bioorg. Med. Chem. Lett.* **9**, 2463–2468 (1999).
20. Han, H., Langley, D. R., Rangan, A. & Hurley, L. H. Selective interactions of cationic porphyrins with G-quadruplex structures. *J. Am. Chem. Soc.* **123**, 8902–8913 (2001).
21. Riou, J. F. *et al.* Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands. *Proc. Natl Acad. Sci. USA* **99**, 2672–2677 (2002).
22. Koepfel, F. *et al.* Ethidium derivatives bind to G-quartets, inhibit telomerase and act as fluorescent probes for quadruplexes. *Nucleic Acids Res.* **29**, 1087–1096 (2001).
23. Fedoroff, O. Y. *et al.* NMR-based model of a telomerase-inhibiting compound bound to G-quadruplex DNA. *Biochemistry* **37**, 12367–12374 (1998).
24. Read, M. A. & Neidle, S. Structural characterization of a guanine-quadruplex ligand complex. *Biochemistry* **39**, 13422–13432 (2000).
25. Han, H., Cliff, C. L. & Hurley, L. H. Accelerated assembly of G-quadruplex structures by a small molecule. *Biochemistry* **38**, 6981–6986 (1999).
26. Otwinowski, Z. M. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
27. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
28. Sheldrick, G. M. & Schneider, T. R. SHELX-97: high-resolution refinement. *Methods Enzymol.* **276**, 319–343 (1997).
29. Cambilleau, C. & Horjales, E. TOM: A FRODO subpackage for protein-ligand fitting with interactive energy minimisation. *J. Mol. Graphics* **5**, 175–177 (1987).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements

We are grateful to Cancer Research UK for support of these studies, and various colleagues for discussions, especially M. Read, J. Harrison, G. Chessari and L. Pearl. We thank M. Roe and the ID14 beamline staff for assistance with data collection.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.N. (e-mail: s.neidle@icr.ac.uk). Atomic coordinates for the 12-mer strand and the 22-mer quadruplex have been deposited in the Protein and Nucleic Acid Data Banks (accession codes: 12 mer, 1K8P and UD0016 respectively; 22-mer, 1KF1 and UD0017 respectively).

corrigendum

MEC-2 regulates *C. elegans* DEG/ENaC channels needed for mechanosensation

M. B. Goodman, G. G. Ernst, D. S. Chelur, R. O'Hagan, C. A. Yao & M. Chalfie

Nature **415**, 1039–1042 (2002).

In this Letter, we reported that MEC-4d capped RNA produced currents statistically indistinguishable from those in water-injected oocytes. Those experiments included oocytes cultured in the presence and absence of 300 μM amiloride. On examining additional amiloride-treated oocytes, we find that injecting MEC-4d capped RNA alone can result in a statistically significant amiloride-sensitive current at −85 mV ($P < 0.0001$). Cells expressing MEC-4d produced amiloride-sensitive currents of $-0.22 \pm 0.03 \mu\text{A}$ ($n = 53$, range: -0.007 to $-1.15 \mu\text{A}$), compared to $-0.005 \pm 0.006 \mu\text{A}$ ($n = 19$, range: -0.06 to $+0.06 \mu\text{A}$) for water-injected controls. Although these currents are similar in average size, voltage-dependence and time-dependence to those measured in cells co-expressing MEC-4d and MEC-10d, these results do not alter our conclusion that MEC-4 and MEC-10 form heteromeric channels, because MEC-10d increases the apparent affinity for amiloride of the MEC-4d/MEC-2 channel. Our new finding, however, suggests that MEC-4 forms homomeric channels when expressed alone. We also continue to find that co-expression of human stomatin or MEC-2(114–363) produces a small increase in MEC-4d current ($P < 0.05$). □

Going soft

The best thing since chalk, and updates of the usual software suspects.

Bibliography management

RefWorks

www.refworks.com/ www.csa.com/refworks

For your personal reference database

The new version of the RefWorks' Internet-based bibliographic management software, released in January 2002, provides expanded data organization and dissemination functions. New features include a global edit function for author names, descriptors, and periodicals; user-defined descriptors (keywords) added to multiple records at once; and 'Electronic Journal' now available as a 'New Reference Type'.

ConceptDraw MEDICAL 1.8

CS-Odessa www.conceptdraw.com

Draw your own conclusions

CS-Odessa is now shipping a free update to ConceptDraw MEDICAL version 1.8, compatible with Windows, Mac OS and Mac OS X. This software is designed to produce medical diagrams, flowcharts and charts simply and quickly. New features in version 1.8 of the package include 10 new medical libraries, the facility to save and read in XML format, and the ability to export documents to EPS and PDF formats.

SDMS/LIMS integration

NuGenesis Technologies/LabWare

www.nugenesis.com/ www.labware.com

Joined-up thinking

Integration between the NuGenesis SDMS (scientific data management system) and LabWare's LIMS (laboratory information management system) has been enhanced to give better ways of using instrument reports from within LIMS. The new degree of integration offers a scalable solution for automating the extraction of analytical data from any PC-based instrument report. When data are extracted via the interface, a traceable hyperlink is created and stored with the sample in LabWare LIMS that points back to the original report in SDMS software. Users can produce composite summary reports containing sample and result information from LIMS combined with analytical instrument reports from the NuGenesis SDMS repository.

SigmaPlot 8.0

SPSS Science www.spss.com

Instant access from Excel

The latest version of this well established graphing tool can be directly accessed from Microsoft Excel. Users can select SigmaPlot

from the Excel menu and generate graphs from Excel worksheets, offering the ability to use data derived from in-cell formulas, store data and graphs in one convenient file. In addition to the Excel integration, SigmaPlot 8.0 offers more editing flexibility and ease of use with new toolbar palettes. Other new features include: multiple levels of undo, Microsoft Windows XP and ME support, and an improved Export dialog — supply the desired image size and SigmaPlot calculates the optimal DPI.

MathPlayer Public Beta

Design Science

www.dessci.com

Display mathematics on the web

The best thing in writing equations since chalk, say the makers. A free public beta release of Design Science's new MathPlayer product can now be downloaded from the company's website. MathPlayer is designed to enable high-quality math display in Internet Explorer for Windows. The principle behind the application is to embed MathML into a normal web page. MathML is a standards-based markup language for encoding mathematics using XML, developed by the World Wide Web Consortium's (W3C) Math Working Group (www.w3.org/math).

Chromeleon 6.40

Dionex

www.dionex.com

Take control of chromatography

Release 6.40 of the Chromeleon chromatography management package adds automation support for capillary LC and nano LC instruments from LC Packings, for new Dionex Summit UV detectors, and for the Agilent 1100 MWD. Also new are capabilities to launch external programs upon completing a sample or sequence, summary System Suitability reports and zoom settings for report pages. Chromeleon fully controls over 150 different HPLC and GC instruments from many different suppliers, including Agilent 1100LCs and 6890GCs, and ThermoFinnigan SpectraSystem LCs and AQA mass spectrometers.

Content files

ISI ResearchSoft www.isiresearchsoft.com

Access more content — free

ISI ResearchSoft has released over 500 new content files for the best-selling bibliographic reference management software packages EndNote, ProCite and Reference Manager, which are supplied and supported in the UK,



Download new content from ISI ResearchSoft.

Ireland and Germany by Adept Scientific (www.adeptscience.co.uk). The new and updated content files include connection files to search Internet libraries worldwide, import filters to capture a wide variety of online and CD-ROM data into a personal reference database, and output styles to create properly formatted bibliographies in each journal's style. Users of EndNote, ProCite and Reference Manager can download these new files free of charge from the product websites: www.endnote.co.uk, www.procite.co.uk and www.refman.co.uk.

ZAP

Zymark

www.zymark.com

Automation to order

ZAP — for Zymark Automation Planner — is an interactive software program that allows the user to browse a selection of automation set-up scenarios, and ultimately assemble the configuration that best fits their applications, throughput, capacity, and walkaway time

needs. It is infinitely scalable by allowing users to find a general application and, based on it, decide the level of automation required; this includes everything from a stand-alone workstation to a fully integrated system. ZAP configures more than 600 automation solutions for over 200 applications clustered within 30 groups. Specific application areas include compound logistics, genomics, proteomics, screening, lead optimization, and ADMET.

QuickQuant

ThermoGalactic www.thermogalactic.com

Spectral data analysis

QuickQuant is a new software application for univariate analysis of spectral data, allowing the spectroscopist to use classic methods of band height or band area for quantitative analysis. Designed as an ActiveApp, QuickQuant runs either directly from GRAMS/AI in a workbook page or by itself as a stand-alone application. QuickQuant supports a fully interactive user interface for quantitative predictions and also supports flexible automation features. The application is fully compatible with My Instrument modules for data acquisition, allowing spectra to be quickly predicted as they are measured and collected directly from a spectrometer system.

STN Easy

Chemical Abstracts Service

<http://stneasy.cas.org>

Search for patents online

A new STN Easy 'Patent Tab' provides fast, easy-to-use patent look-up for web users. The Patent Tab feature is designed especially for those who are not patent search experts but can benefit from the vast amount of scientific information contained in patent records. STN Easy is a search tool available on the web, and with Patent Tab users can now search for patents by inventor, patent assignee, patent number, keywords and more, in a patent-centred research environment. Online help messages and examples are available to assist in choosing search terms specific to patents.

PSI-Plot 7

Poly Software International

www.polysoftware.com

Technical plotting and data processing

First released in September 1992 as a DOS program, PSI-Plot is now released as version 7 for Windows. PSI-Plot version 7 incorporates 30 new features including user-defined hot keys, a built-in calculator, expanded data sheet, expanded pre-defined fitting models, nonlinear user-defined fitting with weighting factors and selectable data range, Stine-man interpolation for data smoothing, complex Fourier transform, wavelet transform, dynamic link of data sheet and plot, retrieving data sheet from plot, automatic labelling

on contour levels, and a new easy-to-use legend editor. Several new plotting types have been added to the new version, such as candle stick plot and Pareto chart.

Origin 7.0

OriginLab

www.originlab.com

More power for an established package

The new version of Origin, the Windows-based scientific graphing and data analysis software, is claimed to be the first scientific software to combine presentation-quality graphics, the C language, and elements of the Numerical Algorithms Group (NAG) library in a single package. Interface improvements include modernized text and drawing tools, a Plotting Wizard with Graph Gallery, and a Nonlinear Curve Fitting wizard. Advanced analysis features include Two-Way ANOVA, Survival analysis, and Nonlinear Curve Fitting with Automatic Parameter Initialization. A new Code Builder interface is added to the software to provide a C programming and debugging environment.

STATISTICA Neural Networks

StatSoft

www.statsoft.co.uk

Have you got the nerve?

STATISTICA Neural Networks version 6 is a data mining, exploration and modelling tool for users who need to analyse complex relationships between variables. The popularity of neural network methodology is rapidly growing in a wide range of areas from basic research to data mining applications, business forecasting, risk management and engineering. Applications include uncovering hidden relationships, explaining known patterns and predicting future trends. STATISTICA Neural Networks version 6 offers a comprehensive selection of neural network architectures, algorithms, sampling methods and analytical graphs for regression, classification and time series problems. It also features a set of tools that make the program powerful and very easy to use.

ChemOffice 2002

CambridgeSoft

www.camsoft.com

Desktop chemistry

ChemOffice is well established as a software toolbox for chemists who want to draw chemical structures; visualize and perform molecular models; and create, store and communicate chemical information. ChemOffice 2002 is the latest version, with new features including a customizable graphical user interface, improved linkage to Microsoft Word and Excel and a new rich polymer notation consistent with IUPAC nomenclature and graphical representation.

These notes are compiled in the Nature office from information provided by the manufacturers.

Contacts

Publisher: Fabien Savenay

Editor: Paul Smaglik

Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Laura Pearson (4977)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silke Opstrup (4994)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau

Tel +33 (0) 1 43 20 16 51

Fax +33 (0) 1 43 20 51 52

e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan

Tel +49 89 54 90 57 11/-2

Fax +49 89 54 90 57 20

e-mail: p.phelan@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Director: Ben Crowe

US Sales Manager: Peyton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakutamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.johnson@naturejp.com

Asia-Pacific Advertising Manager:

Kevyn Johnson

naturejobs

A sign of the times?

The careers fair at the Biotechnology Industry Organization's annual meeting in Toronto last week played host to twice as many jobseekers — about 5,000 — as its predecessor in San Diego last year. But does this mean that there are twice as many people unemployed in biotechnology this year? There are no easy answers or indicators, says Ian King, who helped to organize both meetings. In fact, there are several confounding factors.

First is the meeting's location. Although the conference is international, Toronto managed to draw a disproportionate number of Canadian scientists, King says. And those scientists might have had a ready appetite for such fairs, which are *de rigueur* in the United States, but still novel to the north.

The second factor is political circumstances. Last year's meeting occurred in the midst of a series of protests against the World Trade Organization, which also encompassed biotechnology and genetically modified organisms. As a result, the venue for last year's careers fair was changed, and attendees had to be bussed in — which must have reduced the number of attendees in San Diego.

Finally, there are the employers. Although Toronto saw about the same number of organizations represented compared to last year, their composition was different. The San Diego meeting boasted more companies, whereas Toronto drew more government organizations.

If there is any trend to be identified, it might be that of seeking stability — people this year queued up for more established companies and for government positions, rather than for riskier young companies, says King. To that extent, at least, the fair may be a sign of the times.

Paul Smaglik
Naturejobs editor



Contents

POSTDOCS & STUDENTSHIPS

Redressing the balance
between the sexes

p5

MOVERS

Biomedical institute gets its
first director; Sloan-Kettering
prepares for the future; and
more

Back page

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

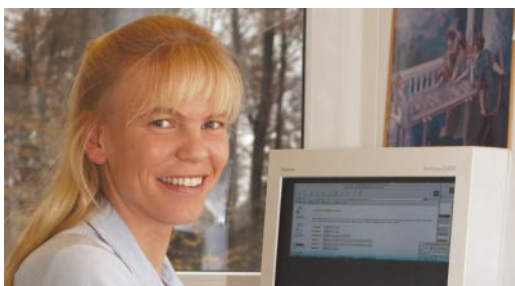
SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

Equal opportunities



Gerlind Wallon:
EMBO plans to help more women scientists come back to work

Women make up about half of all first-degree students in most Western countries. But the proportion goes down as the degree level goes up. By the time women reach the faculty they are under-represented. The proportion of women in academic science varies from a high of 21.5% at lower grades to a mere 5% among full professors, according to a report prepared for the European Commission by the European Technology Assessment Network (ETAN). And female representation is even worse when physics is considered on its own, says a recent report by the International Union for Pure and Applied Physics (IUPAP).

The insecurity of an academic career, being disadvantaged when it comes to promotion and the lack of affordable childcare are among the reasons that many female researchers leave academic research. But it's not all bad news. A small number of funding schemes are aimed at helping them back into work.

The European Molecular Biology Organization (EMBO) announced this February its Restart Fellowship programme for women scientists who have left research to raise a family and would like to return. But this new programme has some limitations. Its leader, Gerlind Wallon, admits that the total of just six fellowships on offer for this year is currently a "sore spot".

FULL OR PART TIME?

The Daphne Jackson Trust in Britain has been offering similar fellowships for the past 10 years, during which time it has helped nearly 100 women scientists and engineers. The Daphne Jackson Fellowship, which is named after the first female physics professor in the United Kingdom, provides two years of funding for women who wish to return to science. In contrast to EMBO's Restart programme, which requires its awardees to return to research full-time, the Daphne Jackson stipends are always funded at half-time levels.

To Althea Wilkinson, now a local project manager at the University of Manchester's Jodrell Bank Observatory, receiving the fellowship was an "absolute lifeline". While she was raising children, her contract as a part-time lecturer was not renewed, and the fellowship enabled her to retrain and obtain her current position.

Raising a family and having a university career in Britain is still a problem for women, according to Yasmin Robson, who was a Daphne Jackson fellow in astrophysics at the University of Oxford. "The part-time nature of the fellowship allows you to do both without the feelings of guilt," she says. Nutritionist Margaret Rayman, who got back into science after a 15-year break and is currently a senior lecturer at the University of Surrey, adds that usually an "enormous proportion of your salary" as a postdoc is spent on childcare.

SHORT TERM

The Marie Heim-Vögtlin (MHV) Fellowships of the Swiss National Science Foundation are for three years, and applicants can choose whether to work full- or part-time. MHV fellow Pia Stieger, a biologist at the University of Berne's Institute of Plant Sciences and a mother of two, was pleased to have this option and is now working 60% of full time.

The current fellowship schemes have a few flaws. Stieger points out that three years is rather a short time to re-enter academic science, get your results published and secure a more permanent position — all while working part-time. Daphne Jackson fellows agree. "Two years is too short to establish a research reputation," says Wilkinson.

Some countries have implemented initiatives that go far beyond funding that is earmarked for female scientists. The Academy of Finland's current equality plan states that "the minority gender is to occupy at least 40% of all research posts". But most countries still shy away from fixing quotas.

Mary Osborn, a group leader at the Max Planck Institute for Biophysical Chemistry who chaired the ETAN Working Group, sees special fellowship programmes as a way of strengthening the position of women in science. In addition, she hopes that more independent positions will be created for young researchers. Osborn's main goal at the moment is to ensure women's representation on advisory boards and appointments committees — which is also a key recommendation of the IUPAP panel. In her opinion, this will be crucial to promoting sexual equality. ■

Jan Schmollinger is a graduate student in Boston.

Web links

European Molecular Biology Organisation
♦ www.embo.org
International Union of Pure and Applied Physics
♦ www.iupap.org
Daphne Jackson Trust
♦ www.daphnejackson.org
Swiss National Science Foundation
♦ www.snf.ch

Transitions

X John Marriott was last month appointed as Britain's Government Chemist by the Department of Trade and Industry. He will be based at LGC, an independent analytical lab in Middlesex, where he has worked since 1998.

X Munich-based GPC Biotech has expanded its clinical-development group in Princeton, New Jersey, by appointing **Edward McNiff** as vice-president, pharmaceutical development; **Thomas McKearn** as vice-president, medical affairs; **Michael Petrone** as vice-president, clinical operations; and **John Slayback** as director for analytical and formulation development/outourcing.

X Ulrich Simon last month became executive vice-president and general manager of the microscopy business group for Carl Zeiss, in Oberkochen, Germany. He has been with Zeiss since 1994, most recently as head of the advanced imaging microscopy division.

X Alison Gadd was this spring appointed as head of regulatory affairs at British Biotech. Gadd joins the Oxford company from Pfizer.

X Elke Jordan, deputy director of the US National Human Genome Research Institute, is to retire.

X William Bonfield and **Michael Redmond** retired this month as non-executive directors of London-based Biocompatibles International.

X Sandy Weinberg, vice-president for entrepreneurship at Muhlenberg College in Allentown, Pennsylvania, has been named as the Linnaeus Chair, a post funded by Amersham Biosciences.

MEDICINE

Although he describes himself as a “dyed-in-the-wool academic”, **Mark Fishman** is leaving his posts at the Massachusetts General Hospital and Harvard Medical School after 25 years to head the new Novartis Institute for Biomedical Research in Cambridge, Massachusetts.

In his new role, he hopes to alter the “sociology of science”, making it easier for scientists in academia and industry to work together and to move back and forth between the sectors. “The new sociology should have the best of academia, biotech and industry,” Fishman says. “It should have the sense that it’s an exciting, creative culture, and also that it’s open, and non-secret.” He suspects that issues relating to intellectual property are the biggest obstacles to overcome in achieving such an environment.

One aspect of the job that appeals to him is its scope. Initially, he will have to hire 400 scientists, although eventually this number could more than double. “The future of biology is going to be large-scale,” Fishman says. One of the institute’s main large-scale efforts will be developing animal models — one of Fishman’s strengths. As chief of cardiology at the Massachusetts General Hospital and professor of medicine at Harvard Medical School, Fishman helped to introduce the zebrafish as a model organism for both cardiovascular development and cardiovascular diseases.

One of the institute’s main personnel needs, Fishman predicts, will be finding scientists who can analyse models once they have been created. “The biologist has a sense of what’s a meaningful pathway, or can simplify or interpret complexity,” he says. Fishman suspects that academic scientists are increasingly willing to work in drug discovery. “There’s a sense that their discoveries could contribute directly to health.” That reason — along with his own acceptance of large-scale biology following the draft completion of the human genome — ultimately drew him to the job.

ETHICS AND GENETICS

As assistant director of the US National Human Genome Research Institute in Bethesda, Maryland, **Kathy Hudson** led the policy, planning, education and communication efforts for the Human Genome Project. But this April, she left government work to head the new Genetics and Public Policy Center. The centre is based in Washington and is affiliated with the Berman Bioethics Institute at Johns Hopkins University. There, Hudson will examine the ethical, scientific and legal issues of genetic testing “along the entire chain of reproductive



Mark Fishman



Kathy Hudson



Thomas Kelly



Robert Wittes

decision-making”. This will include thorny issues such as whether or not to screen newborn children for genetic diseases.

The institute will issue policy options and analyses of different possible decisions rather than offer explicit policy recommendations. Hudson, who enjoyed her broad scope in the Human Genome Project, is pleased to be focusing more intensely on a narrower set of issues with her new post.

CANCER RESEARCH

Two scientists who were chosen to help bridge the bench-to-bedside gap at the Memorial Sloan-Kettering Cancer Center in New York have started their new jobs. Both were drawn to the institute in part because of its president, **Harold Varmus**, the former director of the US National Institutes of Health. But each has taken markedly different career paths to reach a similar point. **Thomas Kelly** joined after 30 years at Johns Hopkins University, whereas **Robert Wittes** switched from academia to government, to industry, and back to government before coming to Sloan-Kettering.

Kelly was chairman of molecular biology and genetics at Johns Hopkins. As chairman of the Sloan-Kettering Institute, he will lead the institute’s basic science effort. He left a “terrific environment” at Johns Hopkins because he “was looking for a new challenge”.

At first, he was reluctant to accept Varmus’ offer, but was convinced after Varmus told him about several major initiatives. These include a 23-storey research building, filling new joint appointments with Rockefeller University and Cornell, and raising the number of principal investigators from 100 to about 150.

Wittes will be physician-in-chief at the cancer centre. For him, his new job is like coming home — he spent his early years at the New York institution. He then joined the National Cancer Institute (NCI), but left to spend two years with drug company Bristol Myers. He returned to the NCI and eventually became its director of extramural science.

Although his experiences are varied, he feels that there are common threads that prepare him for his present post. “For much of the past 20 years I’ve spent time helping to conceptualize and run large national or international programmes either in government or industry,” he says. Wittes expects that his experience, combined with Kelly’s science background and under Varmus’ leadership, will help to turn the promise of contemporary science into medical therapies. “If any place could realize the promise of translating science, this one can,” he says.

CONTACTING US AT MOVERS

Please send any information on appointments or job moves to:
 ◆ naturejobseditor@naturedc.com.